

GI-19745307 264 --- **HWVREKPIIYFKLYRQLPGEKEVAVDDAELN** **QINS**
 ORF84 **WO 2006/078318** **HWVREKPIIYFKLYRQLPGEKEVAVDDAELN** **QINSEGO** PCT/US2005/027239
 GI-28810263 264 **HWVREKPIIYFKLYRQLPGEKEVAVDDAELN** **QINSEGO**
 GI-21909640 150 **HWVREKPIIYFKLYRQLPGEKEVAVDDAELN** **QINSEGO**
 GI-19224141 601 **PHSSVKEVEANKEVTIYNHRETLTFSCKRIWENRDRDORPAKIQVQLLQNGQKMPNQIQEV**

GI-19745307 305 --- **QEISVTWVFNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKKED**
 ORF84 305 **QEISVTWVFNQLVTDERGMAYIYSVREVDK** **NGELLEPKD** **YIKKED**
 GI-28810263 305 **QEISVTWVFNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKKED**
 GI-21909640 191 **QEISVTWVFNQLVTDERGMAYIYSVKEVDK** **NGELLEPKD** **YIKKED**
 GI-19224141 661 **TKDNEWSYHEKDLPKYDANNQBYKYSVEEVNVPDGYKVSYLGNDFNTRETEFVLEQNNF**

GI-19745307 349 **GLTVTNTYVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**
 ORF84 349 **GLTVTNTYVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**
 GI-28810263 349 **GLTVTNTYVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**
 GI-21909640 235 **GLTVTNTYVKPTSG** **HYDIEVTFG** **NGHIDITEDTTPDIVSGENQMK**
 GI-19224141 721 **NLEFGNAETNGQSGSKILDEDTLTSEKGGKIWKNDTABNRPQAIQVQLYADGVAVEGQTK**

GI-19745307 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**
 ORF84 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**
 GI-28810263 394 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**
 GI-21909640 280 **QIEGEDS** **KPIDEVT** **ENNLIEFGKNTMPGEE**
 GI-19224141 781 **FISGSGNEWSFEFKNLKKYNGTGNDIIYSVKEVTVPTGYDVTYSANDLIINTKREVI TOG**

GI-19745307 424 **DGTNSNKYEEVEDSRPVDTL SGLSSEQCGSGDNTI EEDSATHI KFSKRD**
 ORF84 424 **DGTNSNKYEEVEDSRPVDTL SGLSSEQCGSGDNTI EEDSATHI KFSKRD**
 GI-28810263 424 **DGTNSNKYEEVEDSRPVDTL SGLSSEQCGSGDNTI EEDSATHI KFSKRD**
 GI-21909640 310 **DGTNSNKYEEVEDSRPVDTL SGLSSEQCGSGDNTI EEDSATHI KFSKRD**
 GI-19224141 841 **PKLEIEETLPLS GASGGTTIVEDSRPVDTL SGLSSEQCGSGDNTI EEDSATHI KFSKRD**

GI-19745307 473 **IDGKELAGATMELRDS SGNTI STWI SDGQVND FYLMPGKYTFVETAAPDCYEIATAITFT**
 ORF84 473 **IDGKELAGATMELRDS SGNTI STWI SDGQVND FYLMPGKYTFVETAAPDCYEIATAITFT**
 GI-28810263 473 **IDGKELAGATMELRDS SGNTI STWI SDGQVND FYLMPGKYTFVETAAPDCYEIATAITFT**
 GI-21909640 359 **IDGKELAGATMELRDS SGNTI STWI SDGQVND FYLMPGKYTFVETAAPDCYEIATAITFT**
 GI-19224141 901 **IDGKELAGATMELRDS SGNTI STWI SDGQVND FYLMPGKYTFVETAAPDCYEIATAITFT**

GI-19745307 533 **VNEQGOVTVNGKATHGDAHIVW DAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**
 ORF84 533 **VNEQGOVTVNGKATHGDAHIVW DAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**
 GI-28810263 533 **VNEQGOVTVNGKATHGDAHIVW DAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**
 GI-21909640 419 **VNEQGOVTVNGKATHGDAHIVW DAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**
 GI-19224141 961 **VNEQGOVTVNGKATHGDAHIVW DAYKPTKSGQVIDIEEKLPDEQGHSGSTTEIEDSKS**

GI-19745307 593 **SDVIIIGGQC**
 ORF84 593 **SDVIIIGGQC**
 GI-28810263 593 **SDVIIIGGQGEVVDTTEDTQSGMTGHS**
 GI-21909640 479 **SDVIIIGGQGEVVDTTEDTQSGMTGHS** **SGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGMT**
 GI-19224141 1021 **SDLIIGGQGEVVDTTEDTQSGMTGHS**

GI-19745307 602 **QIVETTEDTQ TGMHGDSCCKTEVEDTKLVQSFHFDNK**
 ORF84 602 **QIVETTEDTQ TGMHGDSCCKTEVEDTKLVQSFHFDNK**
 GI-28810263 619 **GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQ TGMHGDSCCKTEVEDTKLVQSFHFDNK**
 GI-21909640 539 **GHS** **GSTTKIEDSKSSDVIIGGQGEVVDTTEDTQ TGMHGDSCCKTEVEDTKLVQSFHFDNK**
 GI-19224141 1047 **GSTTEIEDSKSSDVIIGGQGEVVDTTEDTQ TGMHGDSCCKTEVEDTKLVQSFHFDNK**

GI-19745307 639 **ESESNSEI PKKDKPKSNTSLPATGERQHNMFFIMVTSCSLISSVFVISLKTNRLLSSC**
 ORF84 639 **ESESNSEI PKKDKPKSNTSLPATGERQHNMFFIMVTSCSLISSVFVISLKTNRLLSSC**
 GI-28810263 676 **EFESNSEI PKKDKPKSNTSLPATGERQHNMFFIMVTSCSLISSVFVISLKTNRLLSSC**
 GI-21909640 599 **EPESNSEI PKKDKPKSNTSLPATGERQHNMFFIMVTSCSLISSVFVISLKTNRLLSSC**
 GI-19224141 1104 **EPESNSEI PKKDKPKSNTSLPATGERQHNMFFIMVTSCSLISSVFVISLKTNRLLSSC**

FIGURE 54A

GI-19224WO 2006/078318-MRKNKLLLLATAI LATALGTASLNQNVRAETAGVVSSCOLTKRKSITPCT/US2005/027239

ORF80 1 LBRREKMRKNKLLLLATAI LATALGTASLNQNVRAETAGVVICKSLQVIKMT-YDDEEVLV
GI-21909636 1 MRKNKLLLLATAI LATALGTASLNQNVRAETAGVSENKLVKNTFDSYTDNEVLM
GI-28810259 1 MEREKMRKNKLLLLATAI LATALGTASLNQNVRAETAGVSENKLVKNTFDSYTDNEVLM
GI-19745303 1 MRKNKLLLLATAI LATALGTASLNQNVRAETAGVSDGSLVKKTFPSYTDKVLV
GI-13621428 1 MNLRLHLLLTGAAALTSFAAT-VHGETVVCARLTVTKNLDLVNSN-AL

GI-19224137 56 PNTDYTFVSNPDSAAATCTESN-LPIKPGIAVN-NQDIKVSYSNTDRTSGNEHQVVVDFMK
ORF80 60 PETAEFTLEPDMTASGREGS-LDIRNGIVEGLDKQVTVKYKNTDRTSQNTRLAQDFDFSK
GI-21909636 56 PNADYTFKVEADSTASGRTKDGLEIKPGIVNGLT-EQILSYTWIDRPSKVKSTIEFDFSK
GI-28810259 61 PNADYTFKVEADSTASGRTKDGLEIKPGIVNGLT-EQILSYTWIDRPSKVKSTIEFDFSK
GI-19745303 56 PNADYTFKVEADSTASGRTKDGLEIKPGIVNGLT-EQILSYTWIDRPSKVKSTIEFDFSK
GI-13621428 48 PNTDYTFKVEADSTASGRTKDGLEIKPGIVNGLT-EQILSYTWIDRPSKVKSTIEFDFSK

GI-19224137 114 VTFPSVGIYRYVVTENKGTAE-CVYDDETHLVDVYVGN-NEKCGLEPKYIVSKKEDSA
ORF80 119 VKFPAIGVYRYVMVSEKNDKRD-GITYDTRKIVVDVYVGNKANNBEGFEVLYIVSKEGTSS
GI-21909636 115 VVFPGIGVYRYTVSEKQGDVE-GITYDTRKIVVDVYVGN-KEGGGFEPKIVSKEQGT
GI-28810259 120 VVFPGIGVYRYTVSEKQGDVE-GITYDTRKIVVDVYVGN-KEGGGFEPKIVSKEQGT
GI-19745303 116 VKFPGVGVYRYTVSEVNCNNA-GIAYDSQIVVDVYVGN-REDGGFEAKYIVSKEGQS
GI-13621428 103 VTFEKPVSYYIKVTEEKIDRVPEVSYDITSYTVQVHVLWN-BEQOKPVATYIVGYKEGS-

GI-19224137 171 TNEPTOFNNSFETTSKRTENVTGNTGDKKNAFTFTLTLQFNEYWEASSVVKIEENGQ--
ORF80 178 TKKPIEFTNSIKRTSLKIEKQITGNAQDRKNSFNFTLTLQFSEYKGTGSVVKIEQDCS--
GI-21909636 172 VKKPYNFNNSFATTSKLVKKNVSCNTGELQEFDFTLTLNESTNERNDQTVSLQKCN--
GI-28810259 177 VKKPYNFNNSFATTSKLVKKNVSCNTGELQEFDFTLTLNESTNERNDQTVSLQKCN--
GI-19745303 173 DKRPVLFKNEFDTSKLVKKNVSCNTGELQEFDFTLTLNESTNERNDQTVSLQKCN--
GI-13621428 161 -KVPFOFKNSIDSTTLVKKRVSCITGDRSDENEGTLTKANQYKASEKVMLEKTTKGG

GI-19224137 229 ----TNEVKICEAYKFTLNDSSVILSKLPVGINVYKVEEAERNQCGYITATLNDG--EK
ORF80 236 ----KNDVILCTPKKFTLGHCKSVMLSKLPVGINVYKVEEAERNQCGYITATLNDG--EK
GI-21909636 230 ----KNEVKIGTPEYKFKLNGESTIGLQELPVGELVYKVEEAERNQCGYITATLNDG--EK
GI-28810259 235 ----KNEVKIGTPEYKFKLNGESTIGLQELPVGELVYKVEEAERNQCGYITATLNDG--EK
GI-19745303 231 ----TKKVICEDVFTLNDKESVILSKLPVGINVYKVEEAERNQCGYITATLNDG--EK
GI-13621428 220 QAPVQTEASIDQVHFTLNDKESVILSKLPVGINVYKVEEAERNQCGYITATLNDG--EK

GI-19224137 283 LSTYNLC-OEHRTEKTADEIVVTNRRDQVPTGVVGTLPFAVLSIVAIGGVIYITRRKK
ORF80 292 SSEFTLSTONQRTDESADDEIVVTNRRDQVPTGVVGTLPFAVLSIVAIGGVIYITRRKK
GI-21909636 285 SKMYOLD-MEQRTDESADDEIVVTNRRDQVPTGVVGTLPFAVLSIVAIGGVIYITRRKK
GI-28810259 290 SKMYOLD-MEQRTDESADDEIVVTNRRDQVPTGVVGTLPFAVLSIVAIGGVIYITRRKK
GI-19745303 285 TDCYNLC-DSKRTDSTDEIVVTNRRDQVPTGVVGTLPFAVLSIVAIGGVIYITRRKK
GI-13621428 280 KNLACNSTEGETSLDKDMLIEFTNKKDEVPPTGVVGTLPFAVLSIVAIGGVIYITRRKK

GI-19224137 342 A
ORF80 352 A
GI-21909636 344 A
GI-28810259 349 A
GI-19745303 344 A
GI-13621428 340 A

FIGURE 55

173/487

GI-2190WO 2006/078318

PCT/US2005/027239

GI-28810261	1	MLF	SVVMILTMLAFNOTVLAKDSTV
GI-19224139	1	MLF	SVVMILTMLAFNOTVLAKDSTV
ORF82	1	LLFQ	RVKIFLLTIVLSLSVLFKNNERRLLRKYWKMLF
GI-19745305	1	MRKY	WKMLF

GI-21909638	20	QTSISVENVLERAGDSTPF	SIALESIDAMNTIEEIT	IAGSGKASFSPLTFTTV	GOYTYRV
GI-28810261	26	QTSISVENVLERAGDSTPF	SIALESIDAMNTIEEIT	IAGSGKASFSPLTFTTV	GOYTYRV
GI-19224139	26	QTSISVENVLERAGDSTPF	SIALESIDAMNTIEEIT	IAGSGKASFSPLTFTTV	GOYTYRV
ORF82	61	QTSISVENVLERAGDSTPF	SIALESIDAMNTIDEIT	IAGSGKASFSPLTFTTV	GOYTYRV
GI-19745305	32	QTSISVENVLERAGDST	SIALESIDAMNTIDEIT	IAGSGKASFSPLTFTTV	GOYTYRV

GI-21909638	80	YQKPSQNKDYQADTTV	FDVLVYVTYDEDGTL	VAKVISRRAGDEEKSAIT	FKPKRLVKPIP
GI-28810261	86	YQKPSQNKDYQADTTV	FDVLVYVTYDEDGTL	VAKVISRRAGDEEKSAIT	FKPKRLVKPIP
GI-19224139	86	YQKPSQNKDYQADTTV	FDVLVYVTYDEDGTL	VAKVISRRAGDEEKSAIT	FKPKRLVKPIP
ORF82	121	YQKPSQNKDYQADTTV	FDVLVYVTYDEDGTL	VAKVISRRAGDEEKSAIT	FKPKRLVKPIP
GI-19745305	92	YQKPSQNKDYQADTTV	FDVLVYVTYDEDGTL	VAKVISRRAGDEEKSAIT	FKPKRLVKPIP

GI-21909638	140	PRQPNIPKTPPLPLAGE	VKSLGILSIVLLGLL	VLLYVKKLSRL
GI-28810261	146	PRQPNIPKTPPLPLAGE	VKSLGILSIVLLGLL	VLLYVKKLSRL
GI-19224139	146	PRQPNIPKTPPLPLAGE	VKSLGILSIVLLGLL	VLLYVKKLSRL
ORF82	181	PRQPNIPKTPPLPLAGE	VKSLGILSIVLLGLL	VLLYVKKLSRL
GI-19745305	152	PRQPNIPKTPPLPLAGE	VKSLGILSIVLLGLL	VLLYVKKLSRL

FIGURE 56

PCT/US05/27239 174/487

!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913505.pep from: 1 to: 1036 September 15, 2004 18:46

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences
z-scores computed from opt scores

z-score	obs (=)	exp (*)
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	5	5:====*
48	10	5:====*=====
50	12	4:====*=====
52	4	4:====*
54	6	3:===*====
56	4	3:===*==
58	5	2:==*====
60	3	2:==*==
62	0	1:*
64	1	1:*
66	1	1:*
68	0	1:*
70	1	1:*
72	0	0:
74	0	0:
76	1	0:==
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	1	0:==
98	0	0:

FIGURE 57

PCT/US05/27239

```

100 0 0:
102 0 0:
104 0 0:
106 1 0:=
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=
    
```

Joining threshold: 39, opt. threshold: 27, opt. width: 16, reg.-scaled

The best scores are:

initl initn opt z-sc E(54)..

```

/home/morama/gas/pili/align/gi-50913505.pep Begin: 1 End: 1036
! gi|50913505|ref|YP_059477.1| Collag... 6697 6697 3452.1 9.6e-189
/home/morama/gas/pili/align/gi-19224141.pep Begin: 48 End: 144
! gi|19224141|gb|AAL86412.1|AF447492... 63 100 159 105.9 0.023
/home/morama/gas/pili/align/gi-21909640.pep Begin: 147 End: 449
! gi|21909640|ref|NP_663908.1| protei... 35 35 136 96.3 0.08
/home/morama/gas/pili/align/gi-13621428.pep Begin: 57 End: 318
! gi|13621428|gb|AAK33238.1| hypothet... 33 33 91 75.6 1.1
/home/morama/gas/pili/align/gi-50913506.pep Begin: 33 End: 428
! gi|50913506|ref|YP_059478.1| Fimbri... 70 149 86 71.3 1.9
/home/morama/gas/pili/align/gi-13621432.pep Begin: 14 End: 56
! gi|13621432|gb|AAK33241.1| conserve... 40 65 78 68.0 2.9
/home/morama/gas/pili/align/gi-19745301.pep Begin: 241 End: 466
! gi|19745301|ref|NP_606437.1| putati... 52 52 73 64.8 4.3
/home/morama/gas/pili/align/gas15.pep Begin: 492 End: 739
! GAS15 GAS15 43 68 69 61.4 6.6
/home/morama/gas/pili/align/gi-21909636.pep Begin: 176 End: 298
! gi|21909636|ref|NP_663904.1| conser... 31 31 62 60.8 7.1
/home/morama/gas/pili/align/gi-28810259.pep Begin: 181 End: 303
! gi|28810259|dbj|BAC63197.1| hypothe... 31 31 62 60.7 7.2
/home/morama/gas/pili/align/gi-19224139.pep Begin: 90 End: 143
! gi|19224139|gb|AAL86410.1|AF447492... 43 43 54 58.9 8.9
/home/morama/gas/pili/align/gi-19745305.pep Begin: 96 End: 149
! gi|19745305|ref|NP_606441.1| hypoth... 43 43 54 58.8 9
/home/morama/gas/pili/align/orf82.pep Begin: 125 End: 178
! TRANSLATE of: orf82.seq check: 4296... 43 43 54 58.2 9.6
/home/morama/gas/pili/align/gi-21909638.pep Begin: 84 End: 137
! gi|21909638|ref|NP_663906.1| hypoth... 43 43 52 58.0 9.9
\\End of List
    
```

gi-50913505.pep

/home/morama/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

```

SCORES Initl: 6697 Initn: 6697 Opt: 6697 z-score: 3452.1 E(): 9.6e-189
>>/home/morama/gas/pili/align/gi-50913505.pep (1036 aa)
initn: 6697 initl: 6697 opt: 6697 Z-score: 3452.1 expect(): 9.6e-189
Smith-Waterman score: 6697; 100.0% identity in 1036 aa overlap
(1-1036:1-1036)
    
```

```

10 20 30 40 50 60
gi-50913505. MYSRLKRELIVIVINRKKKYKLIRLMVTVGLIFSQVLVLPIRRRLGLQMIQTQTKVIPQEIIVT
    
```

FIGURE 57A

PCT/US2005/027239

```

gi-50913505. MYSRLKRELIVIVINRKKKYKLIRLMVTVGLIFSQLVLPIRRRLGLQMI STQTKVIPQEIVT
                10      20      30      40      50      60
                70      80      90     100     110     120
gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAI
                70      80      90     100     110     120
gi-50913505. QTETQGTQVVATKQKLESENSSLKVALKRESGFEHNATIDASLDTESQGDNSQRSVTQAI
                70      80      90     100     110     120
                130     140     150     160     170     180
gi-50913505. VTMALELRKQGLSIVDTKIVRIQSSTNQRNDITTTLTFKNGLSLEGASTEANDPNVRVGI
                130     140     150     160     170     180
gi-50913505. VTMALELRKQGLSIVDTKIVRIQSSTNQRNDITTTLTFKNGLSLEGASTEANDPNVRVGI
                130     140     150     160     170     180
                190     200     210     220     230     240
gi-50913505. VNPNDTVQTTITPTIKQDADGKVKNLVFTGRLGKQVIIIVSTTRLKEEQTISLDSYGELVID
                190     200     210     220     230     240
gi-50913505. VNPNDTVQTTITPTIKQDADGKVKNLVFTGRLGKQVIIIVSTTRLKEEQTISLDSYGELVID
                190     200     210     220     230     240
                250     260     270     280     290     300
gi-50913505. GAVGLSQKDRPPYSKPITVNILKPKLSSIESLDSKDFEIVKTI DNLYTWDDQFYLLDFI
                250     260     270     280     290     300
gi-50913505. GAVGLSQKDRPPYSKPITVNILKPKLSSIESLDSKDFEIVKTI DNLYTWDDQFYLLDFI
                250     260     270     280     290     300
                310     320     330     340     350     360
gi-50913505. SKQYEV LKTDYQSAKDSTPQTRDILFGEYTV EPLVMNKGHNNTINIYIRSTRPLGLKPIG
                310     320     330     340     350     360
gi-50913505. SKQYEV LKTDYQSAKDSTPQTRDILFGEYTV EPLVMNKGHNNTINIYIRSTRPLGLKPIG
                310     320     330     340     350     360
                370     380     390     400     410     420
gi-50913505. AAPALIQPRSFRLTPRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDE
                370     380     390     400     410     420
gi-50913505. AAPALIQPRSFRLTPRSTRMKRSAPVEKFEGELEHHKRIDYLGDNQNNPDTTIDDKEDE
                370     380     390     400     410     420
                430     440     450     460     470     480
gi-50913505. HDTSDLYRLYLDMTGKKNPLDILVVVDKSGSMOEGIGSVQRYRYYAQRWDDYYSQWVYHG
                430     440     450     460     470     480
gi-50913505. HDTSDLYRLYLDMTGKKNPLDILVVVDKSGSMOEGIGSVQRYRYYAQRWDDYYSQWVYHG
                430     440     450     460     470     480
                490     500     510     520     530     540
gi-50913505. TFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNGLLQRFVNINPENKL
                490     500     510     520     530     540
gi-50913505. TFDYSSYQGESFNRGQIHYRYRGIVSVSDGIRRDDAVKNSLLGVNGLLQRFVNINPENKL
                490     500     510     520     530     540
                550     560     570     580     590     600
gi-50913505. SVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
                550     560     570     580     590     600
gi-50913505. SVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNSLLDPNTLTALHNN
                550     560     570     580     590     600
                610     620     630     640     650     660
gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRS
                610     620     630     640     650     660
gi-50913505. GTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRS
                610     620     630     640     650     660

```

FIGURE 57B

PCT/US05/27239

```

      670      680      690      700      710      720
gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELE
      |||
gi-50913505. QEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLSGEEHYYGITDTAELE
      670      680      690      700      710      720

      730      740      750      760      770      780
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYDKQPDVLRKSKVNDETEILYQKDQVQEAGK
      |||
gi-50913505. KTLNKIVEDSKLSQLGISDSLSQYVDYDKQPDVLRKSKVNDETEILYQKDQVQEAGK
      730      740      750      760      770      780

      790      800      810      820      830      840
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAYEKYKDNEGRYS
      |||
gi-50913505. DIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAYEKYKDNEGRYS
      790      800      810      820      830      840

      850      860      870      880      890      900
gi-50913505. EMGDSDDYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
      |||
gi-50913505. EMGDSDDYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVD
      850      860      870      880      890      900

      910      920      930      940      950      960
gi-50913505. ADNNQKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGKTYLYETKAKLGYTLP
      |||
gi-50913505. ADNNQKLAGVEFELRKEDKKIVWEKGTGTSNGQLNFKYLQKGKTYLYETKAKLGYTLP
      910      920      930      940      950      960

      970      980      990      1000      1010      1020
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNIYQLPSSGGRGSQIFIIVGSMTA
      |||
gi-50913505. ENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNIYQLPSSGGRGSQIFIIVGSMTA
      970      980      990      1000      1010      1020

      1030
gi-50913505. TVALLFYRRQHRKKQY
      |||
gi-50913505. TVALLFYRRQHRKKQY
      1030

```

gi-50913505.pep
/home/morama/gas/pili/align/gi-19224141.pep
gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

SCORES : Init1: 63 Initn: 100 Opt: 159 z-score: 105.9 E(): 0.023
>>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)
initn: 100 init1: 63 opt: 159 Z-score: 105.9 expect(): 0.023
Smith-Waterman score: 159; 36.7% identity in 98 aa overlap
(895-990:48-144)

```

      870      880      890      900      910      920
gi-50913505. SDASVNYMADGREQKLPYKHPVIQVKTVPITFTKVDADNNQKLAGVEFELRKEDKK-IV
      :| :|:|:| | | :| | :| |
gi-19224141. FILGLLLVFIGLSGVSVGHAETRNGANKQGSFEIKKVDQNNKPLPGATFSLTSKDGKGT
      20      30      40      50      60      70

      930      940      950      960      970      980

```

FIGURE 57C

```

PCT/US2005/027239
gi-50913505. WEKGTGTSNGQLNFKYLQKQKTYLYETKAKLGYTLPENPWEVAVANNGDIK-VKHPIEG
: |:::| : : || | | | | | | | | | : | : || | |:::| :|
gi-19224141. VQFTSNDKGIIVDAQNLPQPG-TYTLKEETAPDGYDKTSRTWTVVYVYENGYTKLVENPYNG
80 90 100 110 120 130

990 1000 1010 1020 1030
gi-50913505. ELKSKDGSYMIKNIYQIQLPSSGGRGSQIFIIVGSMTATVALLFYRRQHRKKQY
|: || ||
gi-19224141. EIIISKAGSKDVSSSLQLENPKMSVVSQYKTEVSSGAADFYRNHAAAYFKMSFELKQKDKS
140 150 160 170 180 190
    
```

gi-50913505.pep
/home/morama/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315]

SCORES Initl: 35 Initn: 35 Opt: 136 z-score: 96.3 E(): 0.08
>>/home/morama/gas/pili/align/gi-21909640.pep (656 aa)
initn: 35 initl: 35 opt: 136 z-score: 96.3 expect(): 0.08
Smith-Waterman score: 148; 24.5% identity in 339 aa overlap
(686-1005:147-449)

```

660 670 680 690 700 710
gi-50913505. NVFRSQEGSKLAIDFKARYPNLSIYSLGVSKDINS DTASSPVVL----KYLSGEEHY
:| | | : : |:: : | ||::
gi-21909640. TNLQAVISVEPVIESLPWTS LKPIAQKDITAKKIWVDAPKEKPIIYFKLYRQLPGEKEV-
120 130 140 150 160 170

720 730 740 750 760 770
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDLSQYVDYDKQPDVLVTRKSKVNDETEILYQ
:| |||:: ::: :::: : :| | |:: : : : :|: : |:|
gi-21909640. -AVDDAELKQ-----INSEGGQEI SVTWT-NQLVT--DEKGMAYIY SVKEVDKNGELLEP
180 190 200 210 220

780 790 800 810 820 830
gi-50913505. KDQVQ-EAGKDIIDKVVFPKTT SQPKGKVLTFKSDYKVDDEYTYTLSFNVKASDEAYE
|| : : | | : : | | || : : : || : : : | | : : : : :
gi-21909640. KDYIKKEDGLTVENTYV---KPTSG-HYDIEVTFGNGH-ID--ITEDTPDIVSGEN---
230 240 250 260 270

840 850 860 870
gi-50913505. KYKDNEGRYSEMGDSDTDYGTNQTSSGKGLPSNSDA--SVNY--MADGR-----
::|: ||: |: | |: :| :|: :|: |: | :| :|:
gi-21909640. QMKQIEGEDSKPIDEVTE--NNLIEFGKNTMPGEEDGTNSNKYEVEDSRPVDTL SGLSS
280 290 300 310 320 330

880 890 900 910 920 930
gi-50913505. EQKLPYKHPVIQVKTVPITFTKVDADNNOKLAGVEFELRKEDKKIVWEKGTGTSNGQLN
|| : : : : | | : | : | : || : | : :| | :|:
gi-21909640. EQGQSGDMTIEEDSATHIKF SKRDIDG--KELAGATMELRDS SGKTI---STWISDGQVK
340 350 360 370 380

940 950 960 970 980 990
gi-50913505. FKYLQKQKTYLYETKAKLGYTLPENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNIY
||: || | : || | || |||:| : : : : : : | : | : |
gi-21909640. DFYLM PGK-YTFVETAAPDGY-----EVATAITFTVNEQGQVTVNGKATKGDHIVMV
390 400 410 420 430 440

1000 1010 1020 1030
gi-50913505. KIIYQLPSSGGRGSQIFIIVGSMTATVALLFYRRQHRKKQY
    
```

FIGURE 57D

PCT/US05/27239

|:|::|:

gi-21909640. DAYK-PTKGSQVIDIEEKLPDEQGHSGSTTEIEDSKSSDVIIGGQGEVVDTTEDTQSGM
450 460 470 480 490 500

gi-50913505.pep
/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 33 Initn: 33 Opt: 91 z-score: 75.6 E(): 1.1
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 33 init1: 33 opt: 91 Z-score: 75.6 expect(): 1.1
Smith-Waterman score: 95; 19.9% identity in 271 aa overlap
(568-819:57-318)

540 550 560 570 580 590
gi-50913505. NKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNRSRDAELLKGWSTNSLLDPNTLTAL
:|: : :|:: :|| : |: : | |
gi-13621428. VNGAKLTVTKNLDLVNSNALIPNTDFTFKIEPDTFVNEDGNKFKGVALNTPMTKVITYTNS
30 40 50 60 70 80

600 610 620 630 640 650
gi-50913505. HNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYR-----SGNGS
::|:| ::| : :|: | :| : | || : :| : : :
gi-13621428. DKGGSNTKTAEFDFSEVTFEKGVYVYKVTTEEKIDKVPGVSYDTTSYTVQVHVLWNEEQQ
90 100 110 120 130 140

660 670 680 690 700
gi-50913505. SNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTAS-SSPVVLKYLSGEEH
: : : :| ||| : | :|| :|: :| |:| ::: : :| : : :
gi-13621428. KPVATYIVGYKEGSKVPI-QFKN---SLDSTLTVKKKVSQTGGDRSKDFNFGLTLKANQ
150 160 170 180 190 200

710 720 730 740 750
gi-50913505. YYGITDTAELEKTLN----KIVEDSKLSQL---GISDSLSQYVDYDQKQPDVLT----R
|| :: : :||| : : ::::| | :|: | | :|| :
gi-13621428. YYKASEKVMIEKTTKGGQAPVQTEASIDQLYHFTLKDGESIKVTNLPVGVVDYVVTEDDYK
210 220 230 240 250 260

760 770 780 790 800 810
gi-50913505. KSKVNDETEILYQKQVQVEAGKDEIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTL
: | : :|: | :|: : : :| | : :|: | :|: | :|:
gi-13621428. SEKYYTNVEVSPQDGAVKNIAGNSTEQETSTDKDMT-----ITFTNKKDFEVPTGVAMTV
270 280 290 300 310

820 830 840 850 860 870
gi-50913505. SENVKASDEAYEKYKDNENGRYSEMGDSDTYGTFNQTSSGKGGGLPSNSDASVNYMADGREQ
:
gi-13621428. APYIALGIVAVGGALYFVKKKNA
320 330 340

gi-50913505.pep
/home/morama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes
MGAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 71.3 E(): 1.9
>>/home/morama/gas/pili/align/gi-50913506.pep (556 aa)

FIGURE 57E

initn: 149 initl: 70 opt: 86 Z-score: 71.3 expect(): 1.9
Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
(503-966:33-428)

```

      480      490      500      510      520
gi-50913505. YSQWVYHGTFDYSSYQGESFNRGQIHRYRGIIVSVSDGIRRDDAVKNSLLGVNG-----L
      |:: || :: :||::: :: :::| |
gi-50913506. NRRETVREKILITAKKIMLACLAILAVVGLGMTRVS-ALSKDDTAQLKITNIEGGPTVTL
      10      20      30      40      50      60.

      530      540      550      560      570      580
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNRDAELLKGWSTNS
      : :: | | |:: | | : :| : : :| : :: :| :||::
gi-50913506. YKIGEGVYNTNGDSFINFK----YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN
      70      80      90      100      110

      590      600      610      620      630      640
gi-50913505. LLDENTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSG
      : | :: : |:: : ||| : ||| |::: : ||| : :
gi-50913506. VSISNGTATYNARGASVYIALLTGAT-----DGRTYNPILLAAS----YNGEGNLVTK
      120      130      140      150      160

      650      660      670      680      690      700
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDFKARYPNLSIYSLGVSKDINSDTASSSPVVLKYLKSGE
      | :| ::| :| :::| |; |::: | : |::| :| : | :| :|
gi-50913506. NIDS--KSNYLYGQ--TSVA----KSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTPE
      170      180      190      200      210

      710      720      730      740      750      760
gi-50913505. EHYYGITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYDQKQPDVLVTRKSKVNDETE
      | | :::| | :||::| : : : : : : :| |::| :|
gi-50913506. LPSY-----TKEAVNKTVY-----VSDNMSEGLTF--NFNSLTVEWKGKMANITE
      220      230      240      250      260

      770      780      790      800      810      820
gi-50913505. ILYQKDQVQEAGKDIIDKVVFPTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDE
      | |:::| :| :| : :| : : :| :
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
      270      280      290

      830      840      850      860      870      880
gi-50913505. AYEKYKDNEGRYSEMGSDDTDYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPEVI
      :|: :|:: :| | : : : :| :| : : :| :| :|
gi-50913506. SYKAVVNNKAI VGEENPNKAEFFYSNNPTKCN TYDNLDKPKDK-GNGITSKEDSK----
      300      310      320      330      340      350

      890      900      910      920      930      940
gi-50913505. QVKTVPITFTKVDADNNQKKLAGVEFELRKEDKKIVWEKGTTFGSNGQLNFKYLQKGTKYY
      | | |::| ||| : : : | | : : : :| :|| : : :| |
gi-50913506. IVYTYQIAERKVDV-SKTPILGALFGVYDTSNKLI-DIVTTNKNGYAI STQVSSGK-YK
      360      370      380      390      400

      950      960      970      980      990      1000
gi-50913505. LYETKAKLGYTL PENPWEVAVANNGDIKVKHPIEGELKSKDGSYMIKNYKIYQLPSSGGR
      : | || ||:| : : :| : :
gi-50913506. IKELKAPKGYSLNTETYEITANWVTATVKTSA NSKSTTYTSDKNKATDNSEQVGLKNGI
      410      420      430      440      450      460

```

gi-50913505.pep
/home/morama/gas/pili/align/gi-13621432.pep

gi|13621432|gb|AAK33241.1| conserved hypothetical protein [Streptococcus pyogene

FIGURE 57F

PCT/US2005/027239/81/487

sl

SCORES Init1: 40 Initn: 65 Opt: 78 z-score: 68.0 E(): 2.9
 >>/home/morama/gas/pili/align/gi-13621432.pep (450 aa)
 initn: 65 init1: 40 opt: 78 Z-score: 68.0 expect(): 2.9
 Smith-Waterman score: 78; 37.0% identity in 46 aa overlap
 (368-411:14-56)

```

      340      350      360      370      380      390
gi-50913505. KGHNNTINIYIRSTRLPLGLKPIGAAPALIQPRSFRLT'PRSTRMK--RSAPVEKFEGELE
                               |  :::| |  ::| |  |  :  :|  ||
gi-13621432.      MTRTNYQKKRMTCPVETEDITYRRKKIKGRRQAILAQFEPELV
                               10      20      30      40

      400      410      420      430      440      450
gi-50913505. HHKRIDYLGDNQNNPDTTIDDKEDHDTSDLYRLYLDMTGKKNPLDIEVVVDKSGSMQEG
      ||: |  ||: :  ||
gi-13621432. HHELI---GDSCTCPDCHGTLTEIGSVVQRQELVFIPAQLKRINHVQHAYKQTCSDNSL
      50      60      70      80      90      100
  
```

gi-50913505.pep
 /home/morama/gas/pili/align/gi-19745301.pep

gi|19745301|ref|NP_606437.1| putative collagen binding protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 52 Initn: 52 Opt: 73 z-score: 64.8 E(): 4.3
 >>/home/morama/gas/pili/align/gi-19745301.pep (524 aa)
 initn: 52 init1: 52 opt: 73 Z-score: 64.8 expect(): 4.3
 Smith-Waterman score: 95; 23.7% identity in 245 aa overlap
 (759-989:241-466)

```

      730      740      750      760      770      780
gi-50913505. DSKLSQLGISDLSQYVDYDKQPDVLVTRKSKVNDTEILYQKDQVQEAG-KDIIDKVV
                               :::|  |  :|  |  :|  |::
gi-19745301. ETIDPDFNEGKEIKYTHILGADLFSYANNPRASTNDE--LLSQVKKVLEKGYRD--DSTT
      220      230      240      250      260

      790      800      810      820      830      840
gi-50913505. FTPKTTSQPKGKVTLT---FKSDYKVDD--EYTYTLSFNVKASDEAYEKYKDNEGRYSEM
      ::: |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-19745301. YANLTSVEFRAATQLAIYYFTDSVDLDNLADYHGFALTEALNATKEIVAYAEADRANLP
      270      280      290      300      310      320

      850      860      870      880      890
gi-50913505. GDSDTDY---GTNQTSS--GKGGLPSNSDASVNYMADGREQKLPYKHPVIQVKTVPIFT
      :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-19745301. NISNLDFYVPNSNKYQSLIGTQYHP-ESLVDIIRMEDKQAPIIPITHKLTISKTVTGTI-
      330      340      350      360      370      380

      900      910      920      930      940      950
gi-50913505. KVDADNNQKKLAGVEFELRKEDKKIVWEKGT'TGSN-GQLNFKYLQKGGK-TYYLYETKAKL
      ||  :|  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-19745301. ---AD--KKKEFNFEIHLKSSDQAI--SGTYPTNSGELT---VTDGKATFTLKDGESLI
      390      400      410      420      430

      960      970      980      990      1000      1010
gi-50913505. GYTLPEN-PWEVAVANNNGDIKVKHPIEGELKSKDGSYMKNYKIYQLPSSGGRGSQIFII
      ||  :  :  :  :  :  :  :  :  :  :  :  :  :
gi-19745301. VEGLPSGYSYEITETGASDYEVS--VNGK-NAPDGKATKASVKEDETITFENRKDLVPPT
  
```

FIGURE 57G

PCT/US05/27239

440 450 460 470 480 490

1020 1030

gi-50913505. VGSMTATVALLFYRRQHRKKQY

gi-19745301. GLTTDGAIYLWLLLLLVLLGLWVWLIQRKGLKND
500 510 520

gi-50913505.pep
/home/morama/gas/pili/align/gas15.pep

GAS15 GAS15

SCORES Init1: 43 Initn: 68 Opt: 69 z-score: 61.4 E(): 6.6
>>/home/morama/gas/pili/align/gas15.pep (762 aa)
initn: 68 init1: 43 opt: 69 Z-score: 61.4 expect(): 6.6
Smith-Waterman score: 100; 21.4% identity in 252 aa overlap
(641-873:492-739)

gi-50913505. AKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSGNGSSNDRNNVTRSQ--EGSKLAI
gas15.pep HIAGRDLFKYTVKPRDTPDPTFLKHIKKVIEKGYRE-KGQAIEYSGLTETQLRAATQLAI
470 480 490 500 510 520

gi-50913505. DEF--KARYPNLSIYSLGVSKDINSSTASSPVVLKYLSGEEHYGTDTAELEKTLNKI
gas15.pep YYFTDSAELDKDKLKYHGFGDMNDSTLAVAKILVEY-AQDSNPPQLTDLDFIPNNNKY
530 540 550 560 570

gi-50913505. VEDSKLSQLGISDSLSQYVDYDQKQPDVLVT----RKSQVN---DETEILYQKDQVQEA
gas15.pep --QSLIGTQWHPEDLVDIIRMEDKKEVIPVTHNLTLRKTVTGLAGDRTKDFHFEIELKNN
580 590 600 610 620 630

gi-50913505. GKDIIDKVVFTPKTTSQPK-GKVTLTFKSDYKVDDE-YTYTSLFNVKASDEAYEKYKDNE
gas15.pep KQELLSQTVKTDKTNLEFKDGKATINLKHGESLTLOGLPEGYSYLVKETDSEGYKVKVNS
640 650 660 670 680 690

gi-50913505. GRYSMGDSDDTYGTDNQT----SSGKGLPSNSDASVN-YMADGREQKLPYKHPVIQVKT
gas15.pep QEVANATVSKTGITSDETLAFENNKPEVPTGVDQKINGYLALIVIAGISLGIWGIHTIR
700 710 720 730 740 750

gi-50913505. VPITFTKVDADNNOKLAGVEFELRKEDKKIVWEKGTGNSGQLNFKYLQKGTYYLYET
gas15.pep IRKHD
760

gi-50913505.pep
/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

FIGURE 57H

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.8 E(): 7.1
 >>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
 initn: 31 init1: 31 opt: 62 Z-score: 60.8 expect(): 7.1
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
 (181-305:176-298)

```

                160      170      180      190      200      210
gi-50913505. DITTTLTFKNGLSLEGASTEANDPNVRVGVNPNDFVQTTTPTIKQDADGKVKNLVFTGR
gi-21909636. VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSKVKKNVSGN-----TGE
                150      160      170      180      190
                220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKL
gi-21909636. LQKEFDFTLTLNESTNFKKDQIVSLQK-GNEKFEVKIGTPYKFKLNGESIQLDKLPVGI
                200      210      220      230      240      250
                270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVKTDYQSAKDSTPQTRDI
gi-21909636. TYKVNEMEANK-DGYKTTASLKEGQSKMYQLDMEQKTDESADEIVVTNKRDTQVPTGV
                260      270      280      290      300      310
                330      340      350      360      370      380
gi-50913505. LFGEYTVPEPLVMNKGHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRLTPRSTRMKRS
gi-21909636. VGTLPFAVLSIVAIGGVIIYITKRKKA
                320      330      340

```

gi-50913505.pep
 /home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 31 Initn: 31 Opt: 62 z-score: 60.7 E(): 7.2
 >>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
 initn: 31 init1: 31 opt: 62 Z-score: 60.7 expect(): 7.2
 Smith-Waterman score: 71; 22.9% identity in 131 aa overlap
 (181-305:181-303)

```

                160      170      180      190      200      210
gi-50913505. DITTTLTFKNGLSLEGASTEANDPNVRVGVNPNDFVQTTTPTIKQDADGKVKNLVFTGR
gi-28810259. VDVYVGNKEGGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSKVKKNVSGN-----TGE
                160      170      180      190      200
                220      230      240      250      260
gi-50913505. LGKQ----VIIVSTTRLKEEQTISLDSYGELVIDGAVGLSQKDRPPYSKPITVNILKPKL
gi-28810259. LQKEFDFTLTLNESTNFKKDQIVSLQK-GNEKFEVKIGTPYKFKLNGESIQLDKLPVGI
                210      220      230      240      250      260
                270      280      290      300      310      320
gi-50913505. SSISSLDSKDFEIVKTIIDNLYTWDDQ--FYLLDFISKQYEVKTDYQSAKDSTPQTRDI
gi-28810259. TYKVNEMEANK-DGYKTTASLKEGQSKMYQLDMEQKTDESADEIVVTNKRDTQVPTGV
                270      280      290      300      310      320

```

FIGURE 57I

PCT/US05/27239

```

330      340      350      360      370      380
gi-50913505. LFG EYTV EPLVMNK GHNNTINIYIRSTRPLGLKPIGAAPALIQPRSFRLTPRSTRMKRS
gi-28810259. VGT LAPFAVLSIVAIGGVIIYITKRKKA
330      340

```

```

gi-50913505.pep
/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

```

```

SCORES  Init1: 43  Initn: 43  Opt: 54  z-score: 58.9  E(): 8.9
>>/home/morama/gas/pili/align/gi-19224139.pep (189 aa)
  initn: 43  init1: 43  opt: 54  Z-score: 58.9  expect(): 8.9
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:90-143)

```

```

720      730      740      750      760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL
gi-19224139. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
60      70      80      90      100      110

770      780      790      800      810      820
gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTILSFNVKASDEAY
gi-19224139. VAKVISRRAGDEEKSATTFKPKRLVKPIPPROPNIPTPLPLAGEVKSLLGILSIVLLGL
120      130      140      150      160      170

```

```

gi-50913505.pep
/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS82
32]

```

```

SCORES  Init1: 43  Initn: 43  Opt: 54  z-score: 58.8  E(): 9
>>/home/morama/gas/pili/align/gi-19745305.pep (195 aa)
  initn: 43  init1: 43  opt: 54  Z-score: 58.8  expect(): 9
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:96-149)

```

```

720      730      740      750      760
gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDSLSQYVDYYDKQP--DVLVTRKSKVNDETEIL
gi-19745305. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
70      80      90      100      110      120

770      780      790      800      810      820
gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTILSFNVKASDEAY
gi-19745305. VAKVISRRAGDEEKSATTFKPKRLVKPIPPROPDIPTPLPLAGEVKSLLGILSIVLLGL
130      140      150      160      170      180

```

```

gi-50913505.pep
/home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672
generated symbols 1 to: 224.

```

FIGURE 57J

GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 43 Initn: 43 Opt: 54 z-score: 58.2 E(): 9.6
>>/home/morama/gas/pili/align/orf82.pep (224 aa)
initn: 43 init1: 43 opt: 54 Z-score: 58.2 expect(): 9.6
Smith-Waterman score: 54; 31.6% identity in 57 aa overlap
(742-796:125-178)

gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDLSQYVDYDKQP--DVLVTRKSKVNDETEIL
orf82.pep ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
720 730 740 750 760
100 110 120 130 140 150
770 780 790 800 810 820
gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY
orf82.pep VAKVISRRAGDEEKSAITFKPKRLVKPIPPRQPNIPKTPLPLAGEVKSLLEGILSIVLLGL
160 170 180 190 200 210

gi-50913505.pep
/home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 43 Initn: 43 Opt: 52 z-score: 58.0 E(): 9.9
>>/home/morama/gas/pili/align/gi-21909638.pep (183 aa)
initn: 43 init1: 43 opt: 52 Z-score: 58.0 expect(): 9.9
Smith-Waterman score: 52; 31.6% identity in 57 aa overlap
(742-796:84-137)

gi-50913505. GITDTAELEKTLNKIVEDSKLSQLGISDLSQYVDYDKQP--DVLVTRKSKVNDETEIL
gi-21909638. ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVLVYV---TYDEDGTL
720 730 740 750 760
60 70 80 90 100 110
770 780 790 800 810 820
gi-50913505. YQKDQVQEAGKDIIDKVVFTPKTTSQPKGKVTLTFKSDYKVDDEYTYTLSFNVKASDEAY
gi-21909638. VAKVISRRAGDEEKSAITFKPKWLVKPIPPRQPNIPKTPLPLAGEVKSLLEGILSIVLLGL
120 130 140 150 160 170

! Distributed over 1 thread.
! Start time: Wed Sep 15 18:46:18 2004
! Completion time: Wed Sep 15 18:46:21 2004
! CPU time used:
! Database scan: 0:00:00.1
! Post-scan processing: 0:00:02.2
! Total CPU time: 0:00:02.3
! Output File: gi-50913505.fasta

FIGURE 57K

!!SEQUENCE_LIST 1.0

(Peptide) FASTA of: gi-50913506.pep from: 1 to: 556 September 15, 2004 18:45

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes MGAS10394]

TO: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp
Variable pamfactor used
Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences
z-scores computed from opt scores

z-score	obs (=)	exp (*)
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	0	5: *
46	1	5:= *
48	8	5:====*====
50	24	5:====*=====
52	2	4:== *
54	3	3:==*
56	1	3:= *
58	2	2:=*
60	3	2:=*==
62	0	2: *
64	2	1:*=
66	2	1:*=
68	1	1:*
70	2	1:*=
72	0	0:
74	3	0:===
76	1	0:=
78	0	0:
80	0	0:
82	0	0:
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:

FIGURE 58

PCT/US05/27239

```

100      0      0:
102      0      0:
104      0      0:
106      0      0:
108      0      0:
110      0      0:
112      0      0:
114      0      0:
116      0      0:
118      0      0:
>120     1      0:=
    
```

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are: init1 initn opt z-sc E(55)..

```

/home/morama/gas/pili/align/gi-50913506.pep    Begin: 1  End: 556
! gi|50913506|ref|YP_059478.1| Fimbri... 3454  3454  3454  1016.2  4.7e-53
/home/morama/gas/pili/align/orf84.pep        Begin: 316  End: 567
! TRANSLATE of: orf84.seq check: 7868...    57     83    135    75.1    1.2
/home/morama/gas/pili/align/gi-19745307.pep    Begin: 316  End: 567
! gi|19745307|ref|NP_606443.1| protei...    57     83    135    75.1    1.2
/home/morama/gas/pili/align/gi-21909640.pep    Begin: 202  End: 524
! gi|21909640|ref|NP_663908.1| protei...    56     81    134    75.0    1.2
/home/morama/gas/pili/align/gi-28810263.pep    Begin: 316  End: 638
! gi|28810263|dbj|BAC63201.1| protein...    56     82    134    74.7    1.3
/home/morama/gas/pili/align/orf80.pep        Begin: 49  End: 352
! TRANSLATE of: orf80.seq check: 9824...    45     69    113    70.8    2.1
/home/morama/gas/pili/align/gi-19224137.pep    Begin: 25  End: 342
! gi|19224137|gb|AAL86408.1|AF447492_...    45     69    109    69.8    2.4
/home/morama/gas/pili/align/gi-19224141.pep    Begin: 277  End: 645
! gi|19224141|gb|AAL86412.1|AF447492_...    73     73    118    68.9    2.7
/home/morama/gas/pili/align/gi-21909636.pep    Begin: 44  End: 344
! gi|21909636|ref|NP_663904.1| conser...    45     98     96    66.1    3.8
/home/morama/gas/pili/align/gi-28810259.pep    Begin: 49  End: 349
! gi|28810259|dbj|BAC63197.1| hypothe...    45     98     96    66.0    3.8
/home/morama/gas/pili/align/gas15.pep        Begin: 222  End: 470
! GAS15 GAS15                                42     68     96    63.8    5
/home/morama/gas/pili/align/gi-13621428.pep    Begin: 17  End: 340
! gi|13621428|gb|AAK33238.1| hypothet...    41     41     87    63.6    5.2
/home/morama/gas/pili/align/gi-19224135.pep    Begin: 193  End: 462
! gi|19224135|gb|AAL86406.1|AF447492_...    41     41     86    61.0    7
/home/morama/gas/pili/align/gi-50913505.pep    Begin: 503  End: 966
! gi|50913505|ref|YP_059477.1| Collag...    70    149     86    60.1    7.8
/home/morama/gas/pili/align/gi-13621430.pep    Begin: 60  End: 143
! gi|13621430|gb|AAK33240.1| hypothet...    43     67     67    59.2    8.7
/home/morama/gas/pili/align/gi-19745303.pep    Begin: 44  End: 344
! gi|19745303|ref|NP_606439.1| hypoth...    51    106     69    58.4    9.5
\\End of List
    
```

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-50913506.pep

gi|50913506|ref|YP_059478.1| Fimbrial structural subunit [Streptococcus pyogenes
MGAS10394]
    
```

```

SCORES    Init1: 3454   Initn: 3454   Opt: 3454   z-score: 1016.2   E(): 4.7e-53
>>/home/morama/gas/pili/align/gi-50913506.pep                               (556 aa)
   initn: 3454   init1: 3454   opt: 3454   Z-score: 1016.2   expect(): 4.7e-53
Smith-Waterman score: 3454;    100.0% identity in 556 aa overlap
    
```

FIGURE 58A

PCT/US05/27239

(1-556:1-556)

```

      10      20      30      40      50      60
gi-50913506. MTNRRETVREKILITAKKMLACLAILAVVGLGMTRVSALS KDDTAQLKITNIEGGPTVT
|||||
gi-50913506. MTNRRETVREKILITAKKMLACLAILAVVGLGMTRVSALS KDDTAQLKITNIEGGPTVT
      10      20      30      40      50      60

      70      80      90     100     110     120
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSIS
|||||
gi-50913506. LYKIGEGVYNTNGDSFINFKYAEGVSLTETGPTSQEITTIANGINTGKIKPFSTENVSIS
      70      80      90     100     110     120

      130     140     150     160     170     180
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYN GEGNLVTKNIDSKSNYLYGQTSV
|||||
gi-50913506. NGTATYNARGASVYIALLTGATDGRTYNPILLAASYN GEGNLVTKNIDSKSNYLYGQTSV
      130     140     150     160     170     180

      190     200     210     220     230     240
gi-50913506. AKSSLPSITKKVTGTIDDVNKKTSLGSVLSYSLTFELPSYTK EAVNKTVYVSDNMSEGL
|||||
gi-50913506. AKSSLPSITKKVTGTIDDVNKKTSLGSVLSYSLTFELPSYTK EAVNKTVYVSDNMSEGL
      190     200     210     220     230     240

      250     260     270     280     290     300
gi-50913506. TFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYK
|||||
gi-50913506. TFNFNSLTVEWKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYK
      250     260     270     280     290     300

      310     320     330     340     350     360
gi-50913506. AVVNNKAIVGEEGNPNKAEFFYSNNPTKGNFYDNLDKKPKDKN GITSKEDSKIVYTYQIA
|||||
gi-50913506. AVVNNKAIVGEEGNPNKAEFFYSNNPTKGNFYDNLDKKPKDKN GITSKEDSKIVYTYQIA
      310     320     330     340     350     360

      370     380     390     400     410     420
gi-50913506. FRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTFNKNGYAI STQVSSGKYKIKELKAPKYS
|||||
gi-50913506. FRKVDSVSKTPLIGAIFGVYDTSNKLIDIVTFNKNGYAI STQVSSGKYKIKELKAPKYS
      370     380     390     400     410     420

      430     440     450     460     470     480
gi-50913506. LNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSE QVWLNKNGIFYSIDSRPTGN
|||||
gi-50913506. LNTETYEITANWVTATVKTSANSKSTTYTSDKNKATDNSE QVWLNKNGIFYSIDSRPTGN
      430     440     450     460     470     480

      490     500     510     520     530     540
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGEL PSTGSIGTYLFAIGSA
|||||
gi-50913506. DVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGEL PSTGSIGTYLFAIGSA
      490     500     510     520     530     540

      550
gi-50913506. AMIGAIGIYIVKRRKA
|||||
gi-50913506. AMIGAIGIYIVKRRKA
      550

```

FIGURE 58B

189/487

PCT/US05/27239
gi-50913506.pep
/home/morama/gas/pili/align/orf84.pep

TRANSLATE of: orf84.seq check: 7868 from: 1 to: 2088
generated symbols 1 to: 696.
GETSEQ from morama, September 13, 2004 17:07.

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
>>/home/morama/gas/pili/align/orf84.pep (696 aa)
initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
(232-462:316-567)

	210	220	230	240	250
gi-50913506.	KTTSLGSVLSYSLTFELPSYTKAEVNTVYVSDNMSEGLTFNFNSLTVIEWKGMAN----				
				: : : ::: :: : : :: :	
orf84.pep	EKEVAVDDAELKQINSEGOQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY				
	290	300	310	320	330 340
	260	270	280	290	300 310
gi-50913506.	ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNISYKAVVNNKAIVGEE				
	:: :: : ::: ::: : : : : :: : :				
orf84.pep	IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNGHIDITEDTTPDI--VSGENQMKQIEGED				
	350	360	370	380	390
	320	330	340	350	
gi-50913506.	GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ				
	:: :: :: ::: :: : :: ::: :: : :				
orf84.pep	SKPIDEVTENNLIIEFGKNIMPGEEDGINSNKYEEVEDSRPVDTLGSLSEQGGSGDMTIE				
	400	410	420	430	440 450
	360	370	380	390	400 410
gi-50913506.	-----IAFRKVDVSVKTPPLIGAFVGYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE				
	:: : : :: : : : : : :				
orf84.pep	EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE				
	460	470	480	490	500 510
	420	430	440	450	460 470
gi-50913506.	LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGLKNGIF				
	: : :: : :: ::: :: :::: :				
orf84.pep	TAAPDGYEIAATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKSGQVIDIEEKL				
	520	530	540	550	560 570
	480	490	500	510	520 530
gi-50913506.	YSIDSRPTGNDVKEAYIESTKALTDGTFESKSNESGTVLLETDPNTKLGELPSTGSG				
	DEQGHSGSTFEIEDSKSSDVIIGGQGQIVETTEDTQTGMHGDGSGCKTEVEDTKLVQSFHF				
orf84.pep	580	590	600	610	620 630

gi-50913506.pep
/home/morama/gas/pili/align/gi-19745307.pep

gi|19745307|ref|NP_606443.1| protein F2-like protein [Streptococcus pyogenes MGA S8232]

SCORES Init1: 57 Initn: 83 Opt: 135 z-score: 75.1 E(): 1.2
>>/home/morama/gas/pili/align/gi-19745307.pep (696 aa)
initn: 83 init1: 57 opt: 135 Z-score: 75.1 expect(): 1.2
Smith-Waterman score: 146; 24.4% identity in 262 aa overlap
(232-462:316-567)

FIGURE 58C

```

                210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKKAVNKTIVYVSDNMSEGLTFNFNLSLTVIEWKGMAN----
                |::: |::: :: |::: |:::
gi-19745307. EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
                290      300      310      320      330      340

                260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNI SYKAVVNNKAI V GEE
                |::: |::: |::: |::: |::: |:::
gi-19745307. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
                350      360      370      380      390

                320      330      340      350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNGITSKEDSKIVYTYQ
                ::| |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |:: |::
gi-19745307. SKPIDEVTENNLIIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTL SGLSSEQQQSGDMTIE
                400      410      420      430      440      450

                360      370      380      390      400      410
gi-50913506. -----IAFRKVDSVSKTPLIGAI FGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
                | | | | :: | | | | : |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
gi-19745307. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMFGKYTFVE
                460      470      480      490      500      510

                420      430      440      450      460      470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
                || || : | |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |:::
gi-19745307. TAAPDGYEIAATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKSGSQVIDIEEKLP
                520      530      540      550      560      570

                480      490      500      510      520      530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNESGTVLLETDIPNTKLGELPSTGSIG
gi-19745307. DEQGHSGSTTEIEDSKSSDVIIGGQGVIVETTEDTQTGMHGDGCKTEVEDTKLVQSFHF
                580      590      600      610      620      630

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-21909640.pep

gi|21909640|ref|NP_663908.1| protein F2 like fibronectin-binding protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 56 Initn: 81 Opt: 134 z-score: 75.0 E(): 1.2
>>/home/morama/gas/pili/align/gi-21909640.pep (656 aa)
initn: 81 init1: 56 opt: 134 z-score: 75.0 expect(): 1.2
Smith-Waterman score: 156; 23.9% identity in 347 aa overlap
(232-547:202-524)

```

                210      220      230      240      250
gi-50913506. KTTSLGSLVLSYSLTFELPSYTKKAVNKTIVYVSDNMSEGLTFNFNLSLTVIEWKGMAN----
                |::: |::: :: |::: |:::
gi-21909640. EKEVAVDDAELKQINSEGQQEISVTWTNQLVTDE--KGMAYIYSVKEVDKNGELLEPKDY
                180      190      200      210      220

                260      270      280      290      300      310
gi-50913506. ITEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNI SYKAVVNNKAI V GEE
                |::: |::: |::: |::: |::: |:::
gi-21909640. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITEDTTPDI-VSGENQMKQIEGED
                230      240      250      260      270      280

```

FIGURE 58D

PCT/US05/27239

```

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNIGITSKEDSKIVYTYQ
          290          300          310          320          330          340
gi-21909640. SKPIDEVTENNLIIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGSLSEQQSGDMTIE

          360          370          380          390          400          410
gi-50913506. -----IAFRKVDVSVSKTPLIGALFGVYDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKE
          350          360          370          380          390          400
gi-21909640. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDQGVKDFYLMGKYTFVE

          420          430          440          450          460          470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKRPSANSKSTTYTSDKNKATDNSEQVGLKNGIF
          410          420          430          440          450
gi-21909640. TAAPDGYEVATAITFTVNEQQQVTVNGKATKGDHIVMV-DAYKPTKGGSGQV-----

          480          490          500          510          520          530
gi-50913506. YSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGELPSTGSIG
          460          470          480          490          500
gi-21909640. --IDIEEKLPD-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-

          540          550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
          510          520          530          540          550          560
gi-21909640. TEIEDSKSSDVIIGGQGEVVDTEDTQSGMTGHSGSTTKIEDSKSSDVIIVGGQGIIVETT

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-28810263.pep

gi|28810263|dbj|BAC63201.1| protein F2-like protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 56 Initn: 82 Opt: 134 z-score: 74.7 E(): 1.3
>>/home/morama/gas/pili/align/gi-28810263.pep (733 aa)
initn: 82 initl: 56 opt: 134 Z-score: 74.7 expect(): 1.3
Smith-Waterman score: 155; 23.6% identity in 347 aa overlap
(232-547:316-638)

```

          210          220          230          240          250
gi-50913506. KPTSLGSVLSYSLTFELPSYTKKAVNKIVYVSDNMSEGLTFNENSLTVEWKGKMAN-----
          290          300          310          320          330          340
gi-28810263. EKEVAVDDAELKQINSEGOQETISVTWENQLVIDE--KGMAYFYSVKEVDKNGELLEPKDY

          260          270          280          290          300          310
gi-50913506. IPEDGSVMVENTKIGIAKEVNNGFNLSFIY-----DSLESISPNI SYKAVVNNKAIVGEE
          350          360          370          380          390
gi-28810263. IKKEDGLTVTNTYV---KPTSGHYDIEVTFGNHIDITETTPDI-VSGENQMKQIEGED-

          320          330          340          350
gi-50913506. GNP-----NKAEFFYSNNP-----TKGNTYDNL-DKKP-DKGNIGITSKEDSKIVYTYQ
          400          410          420          430          440          450
gi-28810263. SKPIDEVTENNLIIEFGKNTMPGEEDGTNSNKYEEVEDSRPVDTLGSLSEQQSGDMTIE

          360          370          380          390          400          410

```

FIGURE 58E

PCT/US05/27239

```

gi-50913506. -----IAFRKVDVSVKTPPLIGALFGVYDTSNKLIDIVTTNKNKYAISTQVSSGKYKIKE
      | | | | : : | | : : | : | | : | : : | : : : | | | : |
gi-28810263. EDSATHIKFSKRD-IDGKELAGATMELRDSSGKTIS--TWISDGQVKDFYLMPGKYTFVE
      460          470          480          490          500          510

```

```

      420          430          440          450          460          470
gi-50913506. LKAPKGYSLNTE-TYEITAN-WVTATVKTSANSKSTTYTSDKNKATDNSEQVGWLKNGIF
      | | | : | | : : : | | : : : : : : | | | : | |
gi-28810263. TAAPDGYEVATAITFTVNEQGQVTVNGKATKGDHIVMV-DAYKPTKGSQV-----
      520          530          540          550          560

```

```

      480          490          500          510          520          530
gi-50913506. YSIDSRPVGNVKEAYIESTKALTDGTTFSKSNESGTVLLETDPNTKLGELPSTGSIG
      | | : | : : : | | : | : : : | : | : | : | : |
gi-28810263. --IDIEEKLDP-EQGHSGSTTEIEDSKSSDVIIGGQGEVVDTE--DTQSGMTGHSGST-
      570          580          590          600          610          620

```

```

      540          550
gi-50913506. TYLFKAIGSAAMIGAIGIYIVKRRKA
      | : : : | : : : |
gi-28810263. TKIEDSKSSDVIVGGQGVIVETTEDTQTGMHGDSGRKTEVEDTKLVQSFHFDNKEPESNS
      630          640          650          660          670          680

```

gi-50913506.pep
/home/morama/gas/pili/align/orf80.pep

TRANSLATE of: orf80.seq check: 9824 from: 1 to: 1056
generated symbols 1 to: 352.
GETSEQ from morama, September 13, 2004 17:11.

SCORES Initl: 45 Initn: 69 Opt: 113 z-score: 70.8 E(): 2.1
>>/home/morama/gas/pili/align/orf80.pep (352 aa)
initn: 69 initl: 45 opt: 113 Z-score: 70.8 expect(): 2.1
Smith-Waterman score: 123; 22.8% identity in 311 aa overlap
(284-556:49-352)

```

      260          270          280          290          300          310
gi-50913506. KMANITEDGSVMVENTKIGIAKEVNGFNLSFIYDSLESISPNI SYKAVVN-NKAIVGEE
      : : | | : | : | : : : : : : | : |
orf80.pep. ATALGTASLNQNVKAETAGVVVTGKSLQVTKMTYDDEEVLMPETAFTFTIEPDMTASGKE
      20          30          40          50          60          70

```

```

      320          330          340          350          360          370
gi-50913506. GNPEN-KAEFFYSNNPFTKGNFYDNLDRKPKDNGITTSKEDSKIVYTYQIAFRKVDVSVKTP
      | : : | : : : : | | | : | : | : : : | | : : | : :
orf80.pep. GSLDIKNGIVEGLDKQVTVKYNDRSOKTK-TAQDFSKVKFPATGVRYRVMVSEKNDK
      80          90          100          110          120          130

```

```

      380          390          400          410          420
gi-50913506. LIGALFGVYDTSNKLIDIVTTNKNK-----YAISTQ-VSSGKYKIKELKAPKGYSLN
      | : : | | : : : | | : : : | | : : : | | : : : | | :
orf80.pep. KDGI---TYDDKKWTVVDVYVGNKANNEEGFEVLYIVSKEGTSSTKKPLEFTNSIKTTSLK
      140          150          160          170          180          190

```

```

      430          440          450          460
gi-50913506. TETYEITANW-----VTATVKTSANSKSTTYTSDKNKATDNSEQVG-----WLKNGI
      | : : | | : | | : : | | : : : : : : : : : | : |
orf80.pep. IEK-QITGNAGDRKKSFNFTLTLQPSYYKTGSVVKIEQDGSKKDVTIGTPYKFTLGHGK
      200          210          220          230          240          250

```

470 480 490 500 510

FIGURE 58F

PCT/US05/27239

```

gi-50913506. FYSIDSRPTGNDV-----KEYAI-----ESTKALTDGTTFSKSNEGSGTVLLETDI
::: || : |::| | : | : | : | : | : | :
orf80.pep SVMLSKLPIGINYYLSEDEANKDGYTTTATLKEQGKEKSSDFTLSTQNQKTDESADIVV
260 270 280 290 300 310

```

```

gi-50913506. PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
| : : | || : || | : | : || : || : || : ||
orf80.pep TNKRDTQVP-TGVVGTLPFAVLSIVAIGGV-IYITKRKKA
320 330 340 350

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

SCORES Initl: 45 Initn: 69 Opt: 109 z-score: 69.8 E(): 2.4
>>/home/morama/gas/pili/align/gi-19224137.pep (342 aa)
initn: 69 initl: 45 opt: 109 z-score: 69.8 expect(): 2.4
Smith-Waterman score: 169; 26.0% identity in 334 aa overlap
(257-556:25-342)

```

gi-50913506. NKTVYVSDNMSEGLTFNFNLSLTVIEWKKGKMANITEDGSVMVENTKIGIAKEVNGFNLSFI
| : : : | : : : | | : : | ||
gi-19224137. MKKNKLLLATAILATALGTASLNQNVKAETAGVVSSGQLTIKKSITN-FN----
10 20 30 40

```

```

gi-50913506. YDSLESISPNI SYKAVVN-NKAI VGE EGN-PNKA EFFYSNNPTKGNTYDNLDKKPKDKNG
| : | : | : | || : : | : | | | : : | : | : | | | :
gi-19224137. DDTL--LMPKTDYTF SVNPDSAATGTESNLPIKPGIAVNNQDIK-VSYSNTDKTSGKEKQ
50 60 70 80 90 100

```

```

gi-50913506. ITSKEDSKIVYTYQIAFRKVD SVSKTPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQV
:: : | : : : | | : : | | : || || : : | : | : | : : :
gi-19224137. VVV-DFMKVTFPSVGIYRYVV TENK----GTAEGVTYDDTKWLVDVYVGNNEKGGLEPKY
110 120 130 140 150 160

```

```

gi-50913506. SSGKYKIKELKAPKGY--SLNTETYETANWVTATVKTSANSKSTTYTSDKNKATDNS--
: | : | | : | : | : | : | : | : | : | : | : | : |
gi-19224137. IVSKKGD SATKEPIQFNNSFETSLKIEKE-VTGNIGDHKKAFTFTLTLQPN EY YEASSV
170 180 190 200 210 220

```

```

gi-50913506. ---EQVGWLKN---GIFYSI---DSR-----PTGND--VKEAYIE-----STKALTDG
| : | | : : | : | : | : | : | : | : | : | : | : |
gi-19224137. VKIEENGQTKDVKIGEAYKFTLNDSQSVILSKLPGVGINYKVEEAEANQGGYTTTATLKDG
230 240 250 260 270 280

```

```

gi-50913506. TTFKSNEG----SGTVLLETDI PNTKLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVK
: | | | : : | : | : : | || : || | : | : || : || : |
gi-19224137. EKLSTYNLGQEHKTDKTADEIVVTNNRDTQVP-TGVVGTLPFAVLSIVAIGGV-IYITK
290 300 310 320 330

```

gi-50913506. RRKA
| : ||

FIGURE 58G

PCT/US05/27239 194/487

gi-19224137. RKKA
340

gi-50913506.pep
/home/morama/gas/pili/align/gi-19224141.pep

gi|19224141|gb|AAL86412.1|AF447492_9 protein F2 [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 118 z-score: 68.9 E(): 2.7
>>/home/morama/gas/pili/align/gi-19224141.pep (1161 aa)
initn: 73 init1: 73 opt: 118 Z-score: 68.9 expect(): 2.7
Smith-Waterman score: 174; 23.9% identity in 406 aa overlap
(115-483:277-645)

	90	100	110	120	130	140
gi-50913506.	VSLTETGPTSQEITTIANGINFGKIKPFSTENVSISNGTATYNARGASVY--IALLTGAT					
				:	: :::	: :
gi-19224141.	IYTFDYIAGLDKVLQSAELSLFLENKEVLENTSISNFKSTIGGQEITYKGTVNVLYGNE					
	250	260	270	280	290	300
	150	160	170	180	190	
gi-50913506.	DGRTYNPILLAASYNGEGNLVTKNIDSKSNYLYGQTSVAKSSLPSITKKVTG-----T					
	: :	:	:	:	:	:
gi-19224141.	STKESNYITNGLSNVG-GSIESYNTETGEFVWVYVNPVRTNIPYATMNLWGFGRARSNT					
	310	320	330	340	350	360
	200	210	220	230	240	250
gi-50913506.	ID---DVNKKTTSLGSLVLSLTF--ELPSYTKAVNKTVYVSDNMSEGLTFNENSLTVE					
	:	:	:	:	:	:
gi-19224141.	SDLENDANTSSAELGEIQVYEVPEGEKLPSSYGVVDVTKLTLRTD-ITAGLNGGFQ-----					
	370	380	390	400	410	
	260	270	280	290	300	310
gi-50913506.	WKGKMANITEDGSVMVENTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAIVG					
		:	:	:	:	:
gi-19224141.	-----MTKRQRIDFG--NNIQNKAFLIKV--TGKTDQSGKPLVVQSNLAS					
	420	430	440	450	460	
	320	330	340	350	360	
gi-50913506.	EEGNPNKAEFFYSNNPTKGNFY--DNLDKKPKDKNGITSKEDSKIVVYTY-----QIAF					
	: :	: :	: :	: :	: :	: :
gi-19224141.	FRGASEYAAF-----TPVGGNVYFQNEIALSPSKGSGSGKSEETKPSITVANLKRVAQLRF					
	470	480	490	500	510	
	370	380	390	400	410	420
gi-50913506.	RKVDSVSKTPLIGALFGVYDTSNKLIDI-VTTNKNGYATSTQVSSGKYKIKELKAPKQYS					
	: :	: :	: :	: :	: :	: :
gi-19224141.	KKM-STDNVPLPEAAFELRSSNGNSQKLEASSNTQGEVHFKDLTSGTYDLYETKAPKGYQ					
	520	530	540	550	560	570
	430	440	450	460		
gi-50913506.	-----LNTETYEIT-----ANWVT--ATVKTSANSKSTTYTSDKNKATDNSEQVGLKN					
	:	:	:	:	:	:
gi-19224141.	QVTEKLATVTVDTTKPAEEMVTWGSPPHSSVKVEAN-KEVTIVNHKETLTFSGKKI-WEND					
	580	590	600	610	620	630
	470	480	490	500	510	520
gi-50913506.	GIFYSIDSRPTGNDVKEAYIESTKALTDGTTFSKSNEGSGTVLLETDPNTKLGELPSTG					
	:	:	:	:	:	:
gi-19224141.	----RPDQRPAAKIQQVQLLQNGQKMPNQIQEVTKDNDWSYHFKDLPKYDAKNQYKYSVEE					
	640	650	660	670	680	

FIGURE 58H

gi-50913506.pep
/home/morama/gas/pili/align/gi-21909636.pep

gi|21909636|ref|NP_663904.1| conserved hypothetical protein [Streptococcus pyogenes MGAS315]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.1 E(): 3.8
>>/home/morama/gas/pili/align/gi-21909636.pep (344 aa)
initn: 98 init1: 45 opt: 96 Z-score: 66.1 expect(): 3.8
Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
(298-556:44-344)

```

      270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNNGFNLSFIYDSLESISPNI SYKAVVNNKAI VGEEGNPNKAEFFYSNNPT
      :: : ::|::: : |::| ::: :
gi-21909636. ATALGTASLNQNVKAETAGVSENAKLIVKKT FDSYTDNEVLMPKADYTFKVE---ADSTA
      20      30      40      50      60      70

      330      340      350      360
gi-50913506. KGNTYDNLDKKPKDKNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK
      :|:| |:|: || |:| : | ||: | :| | || :| |:|
gi-21909636. SGKTKDGLEIKPGIVNGLTEQIISYTN TDKPDSKVKST-EFDFSKVVFPGIGVYRYTVSE
      80      90      100      110      120

      370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SLNPTETY
      | : | : |||: : |: : ||:| : : : | : : :| | : : | : | :
gi-21909636. KQ--GDVEGITYDTKKWTVDVYVGNKEGGGFEPKFI VSKEQGTDVKKPVPNFNNSFATTSL
      130      140      150      160      170      180

      430      440      450      460
gi-50913506. EITANWVTATVVKTSAN-----SKSTTYTSDK----NKATDNSE-QVGW-----LKNGI
      :: | | : : : :|||: : |: : :|::: : | :| |||
gi-21909636. KVKKNVSGNTIGELQKEFDFTLTLNESTNFKKDQIVSLQKGNEKFEVKIGTPYKFKLKNGE
      190      200      210      220      230      240

      470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTFSKS-NEGSGTVLL-----ETD-----
      ::|: |:| | :|::| || :| :||:| : : :|:|
gi-21909636. SIQLDKLPVGITYKVNEMEANK---DGYKTASLKEGDGQSKMYQLDMEQKTDSEADEIV
      250      260      270      280      290      300

      520      530      540      550
gi-50913506. IPNTKLGELPSTGSI GTYLEKAI GSAAMIGAI GYIVKRRKA
      : | : :| || :|| :| :| :| :| :| :| :| :| :| :| :| :| :|
gi-21909636. VTNKRDTQVP-TGVVGT LAPFAVLSIVAIGGV-IYITKRKKA
      310      320      330      340

```

gi-50913506.pep
/home/morama/gas/pili/align/gi-28810259.pep

gi|28810259|dbj|BAC63197.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 45 Initn: 98 Opt: 96 z-score: 66.0 E(): 3.8
>>/home/morama/gas/pili/align/gi-28810259.pep (349 aa)
initn: 98 init1: 45 opt: 96 Z-score: 66.0 expect(): 3.8
Smith-Waterman score: 181; 25.3% identity in 312 aa overlap
(298-556:49-349)

FIGURE 58I

PCT/US05/27239

```

270      280      290      300      310      320
gi-50913506. NTKIGIAKEVNGFNLSFIYDSLESISPNISYKAVVNNKAIVGEEGNPNKAEFFYSNNPT
                :: : ::|::: :   |::  :::
gi-28810259. ATALGTASLNQNVKAETAGVSENAKLIVKKTFFDSYTDNEVLMPKADYTFKVE---ADSTA
20          30          40          50          60          70

330      340      350      360
gi-50913506. KGNTFYDNLDKKPKDKNGIT-----SKEDSKIVYTYQIAFRKV-----DSVSK
:|:| |:|: ||  ||:|   :| |||: |::| ||   :||:
gi-28810259. SGKTKDGLEIKPGIVNGLTEQIISYTNTPDKPDSKVKST-EFDFSKVVFPGIGVYRYTVSE
80          90          100         110         120         130

370      380      390      400      410      420
gi-50913506. TPLIGAIFGV-YDTSNKLIDIVTTNKNGYAISTQVSSGKYKIKELKAPKGY--SLNTETY
| : |: |||:: :|: : ||:| :: : : | : ::| | :: |:| :
gi-28810259. KQ--GDVEGITYDTKKWTVDVVYVGNKEGGFEPKFIVSKEQGTDVKKPVNFNNSFATTSL
140         150         160         170         180         190

430      440      450      460
gi-50913506. EITANWVTATVKTSAN-----SKSTTYTSDK----NKATDNSE-QVGW-----LKNGI
:::| | : : :   ::||:: :|:   :|::: :| :|   |||
gi-28810259. KVKKNVSGNTGELQKEFDFTLTLNESTNFKKQIVSLQKNEKFEVKIGTPYKFKLKNGE
200         210         220         230         240         250

470      480      490      500      510
gi-50913506. FYSIDSRPTGNDVKEAYIESTKALTDGTTESKS-NEGSGTVLL-----ETD-----
::|: |:| | :|:| ||  :| :|::| :   :|
gi-28810259. SIQLDKLPVGITYKVNEMEANK---DGYKTTASLKEGDGQSKMYQLDMEQKTDESADIV
260         270         280         290         300

520      530      540      550
gi-50913506. IPNTKLGELPSTGSIGTYLFAIGSAAMIGAIGIYIVKRRKA
: | : ::| || :|| | : | : ||:: |||:|:|
gi-28810259. VTNKRDTQVP-TGVVGTLPFAVLSIVAIGGV-IYITKRKKA
310         320         330         340

```

gi-50913506.pep /home/morama/gas/pili/align/gas15.pep

GAS15 GAS15

```

SCORES  Init1: 42  Initn: 68  Opt: 96  z-score: 63.8  E(): 5
>>/home/morama/gas/pili/align/gas15.pep (762 aa)
  initn: 68  init1: 42  opt: 96  z-score: 63.8  expect(): 5
Smith-Waterman score: 96; 23.4% identity in 269 aa overlap
(283-535:222-470)

```

```

260      270      280      290      300
gi-50913506. GKMANITEDGSVMVENTKIGIAKEVNGFNLSFIYDSLES-ISPNISYKA---VVNN--K
                ||:: ::|:: |:|::| |  | ::
gas15.pep    VWYYSNAPISNPDESFKRESESNLVSTSQLSLMRQALKQLIDPNLATKMPKQVPDFQQL
200         210         220         230         240         250

310      320      330      340      350      360
gi-50913506. AIVGEEGNPNKAEFFYSNNPTKGNTFYDNLDKKPKDKNGITSKEDSKIVYTYQIAFRK--V
:| | : :| : |:| : | :   || | :   :: : | : : || :
gas15.pep    SIFESEDKGDKYNKGYQNLLSGGLVPT---KPPTPGDPPMPNPQPQ---TTSVLIRKYAI
260         270         280         290         300

370      380      390      400      410      420

```

FIGURE 58J

PCT/US05/27239

gi-13621428. E-KTTKGGQAPVQT-----EASIDQLYHFTLKDGESIKVTNLPVGVVDYVVTTEDDYKSEKYT
220 230 240 250 260

gi-50913506. T-----DGTT-----FSKSNEGSGTVLLETDPNPKLGLPSTGSIGTYLFKAIGSAA
| |:: | :| | : | | |:: : : : |::| :|
gi-13621428. TNVEVSPQDGAVKNIAGNSTEQETSTDKDMTITFTNKKDFEVPTGVAMTVAPYIALGIVA
270 280 290 300 310 320

550
gi-50913506. MIGAIGIYIVKRRKA
: ||: |::|:::|
gi-13621428. VGGAL--YFVKKKNA
330 340

gi-50913506.pep
/home/morama/gas/pili/align/gi-19224135.pep

gi|19224135|gb|AAL86406.1|AF447492_3 Cpa [Streptococcus pyogenes]

SCORES Init1: 41 Initn: 41 Opt: 86 z-score: 61.0 E(): 7
>>/home/morama/gas/pili/align/gi-19224135.pep (756 aa)
initn: 41 init1: 41 opt: 86 Z-score: 61.0 expect(): 7
Smith-Waterman score: 101; 19.6% identity in 306 aa overlap
(243-535:193-462)

220 230 240 250 260 270
gi-50913506. SLTFELPSYTKAVNKTIVYVSDNMSEGLTFNFNLSLTVIEWKGMANITEDGSVMVENTKIG
: : : : | : : : : || : : | : : | :
gi-19224135. PKNANGYMDKIEPLNAILVTQQAVWYYS DSSYGNIKTLWASEL-----KDGKIDFEQVKL-
170 180 190 200 210

280 290 300 310 320 330
gi-50913506. IAKEVNNGENLSFIYDSLESISPNI SYKAVVNNKAI VGEENPN--KAEFFYSNNPTKGN
: | : : : : | : | | : | : : : | : | : : | : | :
gi-19224135. -MREAYS-----KLISDDLEETS KNKLPQGSKLNIFVPQDKSVQNLLSAEYVPESPAPGQ
220 230 240 250 260 270

340 350 360 370 380
gi-50913506. TYDNLDDKPKDKNGLT SKEDSKIVYTY-QIAFRKVD SVSKTFLIGALFGVYDTSNKLIDI
: : | : : | : | : : | : : | : | : : | : : | :
gi-19224135. S-----PEPP-----VQTKKTSVIRKYAEGDY SKLLEGATLRLTGE--DILDFQEK--V
280 290 300 310

390 400 410 420 430 440
gi-50913506. VTINKNGYALSTQVSSGKYKIKELKAPKGYSLNTET-YETTANWVTATVKT SANSKSTT-
: | : | : : : | : : : | : : : : : : : : : : : : : :
gi-19224135. FQSNGTGEKI--ELSNGT YTLTETSSPDGYKIAEPIKFRV VNKKVFIVQKDG SQVENPNK
320 330 340 350 360 370

450 460 470 480 490
gi-50913506. -----YTS DKNKATDNSEQVG---WLKNGIFYSIDSRPTGNDVKEAYLESTKALTDGTF
| : : : : : : : : : | : : : : : : : : : : : : : :
gi-19224135. EVAEPYSVEAYS DMQDSNYINPETFTPYGKFYAKNKDKSSQVVYCFN-----ADLHSP
380 390 400 410 420 430

500 510 520 530 540 550
gi-50913506. SKSNEGSGTVLLETDPNPKLGLPSTGSIGTYLFKAIGSAA MIGAIGIYIVKRRKA
: | : : | : | : : | : | : | : | : | : | : | : | :
gi-19224135. PESEDGGGTI--DPDISTMK--EVKYTH TAGSDLFKYALRPRDTNPEDFLKHIKKVIEKG
440 450 460 470 480

FIGURE 58L

gi-19224135. YNKKGDSYNGLTETQFRAATQLAIYYFTDSTDLKTLKTYNNGKGYHGFESMDEKTLAVTK
490 500 510 520 530 540

gi-50913506.pep
/home/morama/gas/pili/align/gi-50913505.pep

gi|50913505|ref|YP_059477.1| Collagen adhesion protein [Streptococcus pyogenes M GAS10394]

SCORES Init1: 70 Initn: 149 Opt: 86 z-score: 60.1 E(): 7.8
>>/home/morama/gas/pili/align/gi-50913505.pep (1036 aa)
initn: 149 init1: 70 opt: 86 Z-score: 60.1 expect(): 7.8
Smith-Waterman score: 120; 21.5% identity in 469 aa overlap
(33-428:503-966)

gi-50913506. NRRETVREKILITAKKMLACLAILAVVGLGMTRVS-ALSKDDTAQLKITNIEGGPTVTL
gi-50913505. YSQWVYHGTDFDYSSYQGESFNRGQIHRYRGIIVSVSDGIRRDDAVKNSLLGVNG-L
gi-50913506. YKIGEGVYNTNGDSFINFK----YAEGVSLTETGPTSQEIT-TIANGINTGKIKPFSTEN
gi-50913505. LQRFVNINPENKLSVIGFQGSADYHAGKWYPDQSPRGGFYQPNLNNSRDAELLKGWSTNS
gi-50913506. VSISNGTATYNARGASVYIALLTGAT-----DGRTYNPILLAAS-----YNGEGNLVTK
gi-50913505. LLDPNTLTALHNNGTNYHAALLKAKEILNEVKDDGRRKIMIFISDGVPTFYFGEDGYRSG
gi-50913506. NIDS--KSNLYLQ--TSVA---KSSLPSITKKVTGTIDDVNKKTTSLGSVLSYSLTFE
gi-50913505. NGSSNDRNNVTRSQEGSKLAIDEFKARYPNLSIYSLGVSKDINSDTASSPVLKYLKLSGE
gi-50913506. LPSY-----TKEAVNKIVY-----VSDNMSEGLTF--NFNSLTVWEWKGKMANITE
gi-50913505. EHYYGITDTAELEKTELNKIVEDSKLSQLGLSDLSQYVDYDKOPDVLVTRKSKVNDETE
gi-50913506. -----DGSVMVENTKIGIAKEVNNGFNLSFIYDS--LESISPNI-----
gi-50913505. ILYQKDQVQAEAGKDIIDKVVFTEPKTTSQPKGKVTLTFKSDYKVDDEYTYTLNLSFNKASDE
gi-50913506. SYKAVVNNKAIIVGEEGNPNKAEFFYSNNPTKNTYDNLDKKPKD-GNGITSKEDSK----
gi-50913505. AYEKYKDNEGRYSEMGSDDTYGTNQTSSGKGLPSNSDASVNYMADGREQKLPYKHPVI
gi-50913506. IVYTYQIAFRKVDV--VSKTPLIGAIFGVYDTSNKLII-DIVTTNKNGYAISTQVSSGK-YK

FIGURE 58M


```

gi-19745303  KGKTRBGLDLPKPGVLDGHEHTKTIHYGNSDKTTAKEKSVNFDNFANVKFPGVGVYRYTVSE
      80      90      100      110      120      130
      380      390      400      410      420
gi-50913506. LIGAIFGV-YDTSNKLIDIVT'TNKN--GYAISTQVSS-GKYKIKELKAPKGYSLNTETYE
      : |  | : || : : : | : : : | : : | : | : | : : : | : :
gi-19745303. VNGNKAGIAYDSQQWTFVDVYVVNREDGGFEAKYIVSTEGGQSDKKPVLFKNF-FDTTSLK
      140      150      160      170      180
      430      440      450      460      470
gi-50913506. ITANWVTATVKTSANSKSTTYTSDKNKATDNSEQVGLKNGIF-----YSIDSR---
      : | : || : : : | | | | : : : : | : | : : | | : :
gi-19745303. VTKK-VTGNTEGEHQRSFSFTLLLPNECFEKGQVVNIIQGETKKVVIGEEYSFTLKDKE
      190      200      210      220      230      240
      480      490      500      510
gi-50913506. -----PTG-----NDVKEAYIESTKALTDGTTFSKSNEG-SGTVLLLETD---IPNT
      | : | : | : : : | | : : | | | | : : | | : |
gi-19745303. SVTLSQLPVGIEYKVTEEDVTKDGYKTSATLKDGDVTDGYNLGDSTDKSTDEIVVTNK
      250      260      270      280      290      300
      520      530      540      550
gi-50913506. KLGELPSTGSIGTYLFKAIGSAAMIGAIGIYIVKRRKA
      : : : | | : | | | : | : | : : | | : | |
gi-19745303. RDTQVP-TGVVGTLPFAVLSIVAIGGV-IYITKRKKA
      310      320      330      340

```

```

! Distributed over 1 thread.
! Start time: Wed Sep 15 18:45:54 2004
! Completion time: Wed Sep 15 18:46:02 2004

! CPU time used:
! Database scan: 0:00:00.1
! Post-scan processing: 0:00:01.9
! Total CPU time: 0:00:02.0
! Output File: gi-50913506.fasta

```

FIGURE 580

!!SEQUENCE_LIST 1.0 13621430/27239

(Peptide) FASTA of: gi-13621430.pep from: 1 to: 215 September 15, 2004 18:45

gi|13621430|gb|AAK33240.1| hypothetical protein [Streptococcus pyogenes]

T0: *.pep Sequences: 56 Symbols: 22,803 Word Size: 2

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

Histogram Key:

Each histogram symbol represents 1 search set sequences

z-scores computed from opt scores

z-score	obs (=)	exp (*)
< 20	0	0:
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	0:
32	0	0:
34	0	1:*
36	0	2: *
38	0	3: *
40	0	4: *
42	0	5: *
44	1	5:== *
46	7	5:====*===
48	21	5:====*=====
50	7	5:====*===
52	2	4:=== *
54	4	3:==*==
56	3	3:==*
58	0	2: *
60	4	2:==*==
62	1	2:==*
64	0	1:*
66	0	1:*
68	0	1:*
70	0	1:*
72	0	0:
74	0	0:
76	0	0:
78	0	0:
80	3	0:===
82	2	0:==
84	0	0:
86	0	0:
88	0	0:
90	0	0:
92	0	0:
94	0	0:
96	0	0:
98	0	0:
100	0	0:

FIGURE 59


```

102 PCT/US05/27239
104 0 0:
106 0 0:
108 0 0:
110 0 0:
112 0 0:
114 0 0:
116 0 0:
118 0 0:
>120 1 0:=

```

Joining threshold: 36, opt. threshold: 24, opt. width: 16, reg.-scaled

The best scores are:

	init1	initn	opt	z-sc	E(55)
/home/morama/gas/pili/align/gi-13621430.pep			Begin: 1 End: 215		
! gi 13621430 gb AAK33240.1 hypothet...	1338	1338	233.9	1.8e-09	
/home/morama/gas/pili/align/gi-19745305.pep			Begin: 1 End: 193		
! gi 19745305 ref NP_606441.1 hypoth...	163	243	273	82.2	0.5
/home/morama/gas/pili/align/gi-28810261.pep			Begin: 2 End: 187		
! gi 28810261 dbj BAC63199.1 hypoth...	164	239	268	81.5	0.55
/home/morama/gas/pili/align/gi-19224139.pep			Begin: 2 End: 187		
! gi 19224139 gb AAL86410.1 AF447492...	164	236	265	81.0	0.57
/home/morama/gas/pili/align/orf82.pep			Begin: 30 End: 222		
! TRANSLATE of: orf82.seq check: 4296...	163	235	264	81.0	0.58
/home/morama/gas/pili/align/gi-21909638.pep			Begin: 2 End: 181		
! gi 21909638 ref NP_663906.1 hypoth...	164	239	261	80.5	0.62
/home/morama/gas/pili/align/gi-19745303.pep			Begin: 84 End: 183		
! gi 19745303 ref NP_606439.1 hypoth...	121	121	126	61.4	6.7
/home/morama/gas/pili/align/gi-13621428.pep			Begin: 6 End: 174		
! gi 13621428 gb AAK33238.1 hypothet...	58	86	122	60.9	7.2
/home/morama/gas/pili/align/gi-19224137.pep			Begin: 93 End: 201		
! gi 19224137 gb AAL86408.1 AF447492...	88	88	119	60.4	7.5
/home/morama/gas/pili/align/gi-50913503.pep			Begin: 549 End: 625		
! gi 50913503 ref YP_059475.1 Fibron...	73	73	117	60.4	7.6
/home/morama/gas/pili/align/gi-19224134.pep			Begin: 631 End: 697		
! gi 19224134 gb AAL86405.1 AF447492...	73	73	115	60.1	7.8

gi-13621430.pep

/home/morama/gas/pili/align/gi-13621430.pep

gi|13621430|gb|AAK33240.1|hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 1338 Initn: 1338 Opt: 1338 z-score: 233.9 E(): 1.8e-09
>>/home/morama/gas/pili/align/gi-13621430.pep (215 aa)
initn: 1338 init1: 1338 opt: 1338 z-score: 233.9 expect(): 1.8e-09
Smith-Waterman score: 1338; 100.0% identity in 215 aa overlap
(1-215:1-215).

```

          10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
          10      20      30      40      50      60

          70      80      90     100     110     120
gi-13621430. EALDKESPLPNSVTTSVKGNKGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
gi-13621430. EALDKESPLPNSVTTSVKGNKGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETVYEVV

```

FIGURE 59A

PCT/US2005/027239

```

          90      100      110      120
          130      140      150      160      170      180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQYSEKTPPEPHQPDTEKEKPKKRNIGI
          |||
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQYSEKTPPEPHQPDTEKEKPKKRNIGI
          130      140      150      160      170      180
          190      200      210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
          |||
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
          190      200      210

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-19745305.pep

gi|19745305|ref|NP_606441.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 163 Initn: 243 Opt: 273 z-score: 82.2 E(): 0.5
>>/home/morama/gas/pili/align/gi-19745305.pep (195 aa)
initn: 243 init1: 163 opt: 273 Z-score: 82.2 expect(): 0.5
Smith-Waterman score: 320; 31.9% identity in 213 aa overlap
(1-213:1-193)

```

          10      20      30      40      50      60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
          |::| :::: ::| : : ::| |:: |:: :| | |:: | :: | | ::
gi-19745305. MRKYWKMLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTS-----FSVAL
          10      20      30      40      50
          70      80      90      100      110      120
gi-13621430. EALDKESPLPNSVTTTSVKGNGKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETIVYEVV
          |::| : : : :| : |::| | |:: | |:: |::| : |::| : | |::| :
gi-19745305. ESIDAMKTI-DEIT--IAGSGKASFSPLTFTFVGQYTYRVYQKPSQNKDYQADTTVFDVL
          60      70      80      90      100      110
          130      140      150      160      170      180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSELIFKQYSEKTPPEPHQPDTEKEKPKKRNIGI
          :|| |::| :| ::::| : |::| | : | : | : | : | : | : | :
gi-19745305. VYVYDDED-GTLVAKVISRRAGDEEKSAITFKPKRLVKPIPPRPDI-----PKTP-----
          120      130      140      150      160
          190      200      210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
          || :||: | :: |::|::: ::| : |::|:
gi-19745305. LPLAGEVKSLGLSIVLLGLLVLLYV-KKLSRL
          170      180      190

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-28810261.pep

gi|28810261|dbj|BAC63199.1| hypothetical protein [Streptococcus pyogenes SSI-1]

SCORES Init1: 164 Initn: 239 Opt: 268 z-score: 81.5 E(): 0.55
>>/home/morama/gas/pili/align/gi-28810261.pep (189 aa)
initn: 239 init1: 164 opt: 268 Z-score: 81.5 expect(): 0.55
Smith-Waterman score: 306; 30.6% identity in 206 aa overlap

FIGURE 59B

(8-213:2-187) 1505/27239

```

10 20 30 40 50 60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
gi-28810261. MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
10 20 30 40
70 80 90 100 110 120
gi-13621430. EALDKESPLPNSVTTTSVKGNKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETIVYEVV
gi-28810261. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
50 60 70 80 90 100
130 140 150 160 170 180
gi-13621430. IYVLYNEQSGALETNLVSINKLGETEKSSELIFKQEYSEKTPEPHQPDTEKEKPKKRNGI
gi-28810261. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLKPIPPRQPNI-----PKTP-----
110 120 130 140 150
190 200 210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
gi-28810261. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLKSRL
160 170 180

```

gi-13621430.pep

/home/morama/gas/pili/align/gi-19224139.pep

gi|19224139|gb|AAL86410.1|AF447492_7 unknown [Streptococcus pyogenes]

```

SCORES Init1: 164 Initn: 236 Opt: 265 z-score: 81.0 E(): 0.57
>>/home/morama/gas/pili/align/gi-19224139.pep (189 aa)
initn: 236 init1: 164 opt: 265 Z-score: 81.0 expect(): 0.57
Smith-Waterman score: 303; 30.6% identity in 206 aa overlap
(8-213:2-187)

```

```

10 20 30 40 50 60
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI
gi-19224139. MLFSVVMILTMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL
10 20 30 40
70 80 90 100 110 120
gi-13621430. EALDKESPLPNSVTTTSVKGNKTSFEQLTFSEVGQYHYKIHQLLGKNSQYHYDETIVYEVV
gi-19224139. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
50 60 70 80 90 100
130 140 150 160 170 180
gi-13621430. IYVLYNEQSGALETNLVSINKLGETEKSSELIFKQEYSEKTPEPHQPDTEKEKPKKRNGI
gi-19224139. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKRLVLPPIPPRQPNI-----PKTP-----
110 120 130 140 150
190 200 210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
gi-19224139. LPLAGEVKSLGILSIVLLGLLVLLYV-KKLKSRL
160 170 180

```

FIGURE 59C

PCT/US05/27239

gi-13621430.pep /home/morama/gas/pili/align/orf82.pep

TRANSLATE of: orf82.seq check: 4296 from: 1 to: 672 generated symbols 1 to: 224. GETSEQ from morama, September 13, 2004 17:09.

SCORES Init1: 163 Initn: 235 Opt: 264 z-score: 81.0 E(): 0.58 >>/home/morama/gas/pili/align/orf82.pep (224 aa) initn: 235 init1: 163 opt: 264 Z-score: 81.0 expect(): 0.58 Smith-Waterman score: 304; 30.5% identity in 213 aa overlap (1-213:30-222)

gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTA orf82.pep LLFQRVKIFLLTIVLSLSVLFKNNERRRLLRKYWKMLFSVVMILMMLAFNQTVLAKDSTV ... SINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLPNSVTTSVKGNKTSFEQLTFS orf82.pep QTSISVENVLERAGDSTP-----FSVALESIDAMKTI-DEIT--IAGSGKASFSPLTFT ... EVGQYHYKIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETEKSELIF orf82.pep TVGQYTYRVYQKPSQNKDYQADTTVFDVLVYVYDDED-GTLVAKVISRRAGDEEKSAITF ... KQEYSEKTPPEHPQDPTTEKEKPQKRNGLPSTGEMVSYSALGIVLVATITLYSIYKKL orf82.pep KPKRLVKPIPPROPNI-----PKTP-----LPLAGEVKSLLGILSIVLLGLLVLLYV-KKL ... KTSK orf82.pep KSRL

gi-13621430.pep /home/morama/gas/pili/align/gi-21909638.pep

gi|21909638|ref|NP_663906.1| hypothetical protein [Streptococcus pyogenes MGAS31 5]

SCORES Init1: 164 Initn: 239 Opt: 261 z-score: 80.5 E(): 0.62 >>/home/morama/gas/pili/align/gi-21909638.pep (183 aa) initn: 239 init1: 164 opt: 261 Z-score: 80.5 expect(): 0.62 Smith-Waterman score: 302; 31.5% identity in 200 aa overlap (14-213:2-181)

gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVDVATNKQSSDIDETFMFVI orf82.pep MILMMLAFNQTVLAKDSTVQTSISVENVLERAGDSTP-----FSIAL

FIGURE 59D

PCT/US05/27239

```

              70          80          90          100          110          120
gi-13621430. EALDKESPLPNSVTTSVKNGKTSFEQLTFSEVGOYHYKIHQLLGKNSQYHYDETVYEVV
              |::| : : : : : |::|::| |::| |::| |::| |::| |::| |::| |::|
gi-21909638. ESIDAMKTIEE---ITIAGSGKASFSPLTFTTVGQYTYRVYQKPSQNKDYQADTTVFDVL
              50          60          70          80          90

              130          140          150          160          170          180
gi-13621430. IYVLYNEQSGALETNLVSNNKLGETEKSSELIFKQYSEKTPPEPHQPDTEKEKPKRNGI
              :|| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
gi-21909638. VYVTYDED-GTLVAKVISRRAGDEEKSAITFKPKWLKPIPPROPNI-----PKTP-----
              100          110          120          130          140

              190          200          210
gi-13621430. LPSTGEMVSYVSALGIVLVATITLYSIYKKLKTSK
              || :||: | : : |::|::|::| : : |::|:
gi-21909638. LPLAGEVKSLLGILSIVLLGLLVLLYV-KKLKSRLL
              150          160          170          180

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-19745303.pep

gi|19745303|ref|NP_606439.1| hypothetical protein [Streptococcus pyogenes MGAS8232]

SCORES Init1: 121 Initn: 121 Opt: 126 z-score: 61.4 E(): 6.7
>>/home/morama/gas/pili/align/gi-19745303.pep (344 aa)
initn: 121 init1: 121 opt: 126 z-score: 61.4 expect(): 6.7
Smith-Waterman score: 126; 27.0% identity in 100 aa overlap
(59-155:84-183)

```

              30          40          50          60          70          80
gi-13621430. LTASINIEVINQVDVATNKQSSDIDETFMFVIEALDKESPLP---NSVTTSVKNGK-KTSF
              |::|::|::|::|::|::| |::| |::| |::| |::|
gi-19745303. LMPKADYTFKVEADDNAKGTKDGLDIKPGVIDGLENTKTIHYGNSDKTTAKEKSVNFDF
              60          70          80          90          100          110

              90          100          110          120          130          140
gi-13621430. EQLTFSEVGOYHYKIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNNKLGETE
              :: | || |::| : : : |::| |::| |::| |::| |::| |::|
gi-19745303. ANVKFPGVGVYRYTVSEVNGNKAGLAYDSQQWTVDVYVNVNREDGGFEAKYLVSTEGGQSD
              120          130          140          150          160          170

              150          160          170          180          190          200
gi-13621430. KSELIFKQYSEKTPPEPHQPDTEKEKPKRNGILPSTGEMVSYVSALGIVLVATITLY
              |: :::|::|: : :
gi-19745303. KKPVLFRNFFDITSLKVTKKVTGNTGEHQRSFSFTLLLPNECFEKGQVVNIIQGETKK
              180          190          200          210          220          230

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-13621428.pep

gi|13621428|gb|AAK33238.1| hypothetical protein [Streptococcus pyogenes]

SCORES Init1: 58 Initn: 86 Opt: 122 z-score: 60.9 E(): 7.2
>>/home/morama/gas/pili/align/gi-13621428.pep (340 aa)
initn: 86 init1: 58 opt: 122 z-score: 60.9 expect(): 7.2
Smith-Waterman score: 135; 29.1% identity in 172 aa overlap

FIGURE 59E

(8-15956171) US05/27239

```

      10      20      30      40      50
gi-13621430. MKKSILRILAIGYLLMSFCLLDSVEAENLTASINIEVINQVD-VATNKQSSDIDETFMF-
      :| | | || :|::|::: : :: | ::| | :| : | || :
gi-13621428.  MKLRHLLLTGAALTSFAAT-TVHGETVVNGAKLTVTKNLDLVNSNALIPNTDFTFKIE
      10      20      30      40      50

      60      70      80      90      100
gi-13621430. ----VIEALDK-----ESPLPN-SVTTSVKGNKTSFEQLTFSEV-----GQYHYKI-H
      | | :| :|::: : | :| |:::| : :| ||| | | :| | :
gi-13621428. PDTTVNEDGNKFKGVALNTPMTKVITYTNSDKGGSNTKTAEFDFSEVTFEKPGVYYYKVT
      60      70      80      90      100      110

      110      120      130      140      150      160
gi-13621430. QLLGKNSQYHYDETVEVVIYVLYNE-QSGALETNLVSNKLGETEKSELIEFKOEYSEKTP
      : : -| || | | :|::| | | : : | :| : | | | : | | : : |
gi-13621428. EKIDKVPVGSYDPTTSYTVQVHVLWNEEQOKPVATYIVGYKEGS--KVPIQFKNSLDSTTL
      120      130      140      150      160      170

      170      180      190      200      210
gi-13621430. EPHQPDTTEKEKPKQKRNIGILPSTGEMVSYVSALGIVLVATITLYSIYKCLKTSK
gi-13621428. TVKKKVSQGTGGDRSKDFNGLTLKANQYYKASEKVMIEKTTKGGQAPVQTEASIDQLYHF
      180      190      200      210      220      230

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-19224137.pep

gi|19224137|gb|AAL86408.1|AF447492_5 EftLSL.A [Streptococcus pyogenes]

SCORES Initl: 88 Initn: 88 Opt: 119 z-score: 60.4 E(): 7.5
>>/home/morama/gas/pili/align/gi-19224137.pep (342 aa)
initn: 88 initl: 88 opt: 119 Z-score: 60.4 expect(): 7.5
Smith-Waterman score: 119; 29.7% identity in 111 aa overlap
(72-176:93-201)

```

      50      60      70      80      90
gi-13621430. DVATNKQSSDIDETFMFVIEALDKESPLNSVTTTSVKGNK-----TSFEQLTFSEVQY
      | : : : | : || : : | : || || |
gi-19224137. SVNPDSAATGTESNLPKPGIAVNNQDIKVSYSNTDKTSGKEKQVVVDFMKVTFPSVGIY
      70      80      90      100      110      120

      100      110      120      130      140      150
gi-13621430. HYKIHQLLGKNSQYHYDETVEVVIYVLYNEQSGALETNLVSNKLGETEKSELI-FKOEY
      :| : : | | :| : | :|| | | : : | | : : | | | : :
gi-19224137. RYVVTENKGTAEQVTVDDTKWLVDVYVGNNEKGG-LPKYIVSKKGD SATKEPIQFNNSF
      130      140      150      160      170      180

      160      170      180      190      200      210
gi-13621430. SEKTPPEPHQPDTTEKEKPKQKRNIGILPSTGEMVSYVSALGFVLVATITLYSIYKCLKTSK
      | | : : | : : ||
gi-19224137. -ETTSLKIEKEVTFGNTGDHKKAF TFLTLQPN EYYEASSVVKIEENGQTKDVKIGEAYKF
      190      200      210      220      230      240

```

gi-13621430.pep
/home/morama/gas/pili/align/gi-50913503.pep

gi|50913503|ref|YP_059475.1| Fibronectin-binding protein [Streptococcus pyogenes
MGAS10394]

FIGURE 59F

PCT/US05/27239

SCORES Init1: 73 Initn: 73 Opt: 117 z-score: 60.4 E(): 7.6
 >>/home/morama/gas/pili/align/gi-50913503.pep (627 aa)
 initn: 73 init1: 73 opt: 117 Z-score: 60.4 expect(): 7.6
 Smith-Waterman score: 118; 28.7% identity in 87 aa overlap
 (129-215:549-625)

```

      100      110      120      130      140      150
gi-13621430. KIHQLLGKNSQYHYDETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEK
gi-50913503. IETEDTKEPEVLMGGQSESVEFTKDTQTGMSGFSETATVV-----EDTRPKLVFHFDDNNEP
      520      530      540      550      560      570

      160      170      180      190      200      210
gi-13621430. TPEPHQPDTEKEKPKQKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKCLKTSK
gi-50913503. KVEEN-----REKPTKNITPILPATGDIENVLAFGLILSVLSIFSLLKKNKQSNKKV
      580      590      600      610      620

```

gi-13621430.pep
 /home/morama/gas/pili/align/gi-19224134.pep
 gi|19224134|gb|AAL86405.1|AF447492_2 protein F [Streptococcus pyogenes]

SCORES Init1: 73 Initn: 73 Opt: 115 z-score: 60.1 E(): 7.8
 >>/home/morama/gas/pili/align/gi-19224134.pep (698 aa)
 initn: 73 init1: 73 opt: 115 Z-score: 60.1 expect(): 7.8
 Smith-Waterman score: 115; 27.4% identity in 73 aa overlap
 (143-215:631-697)

```

      120      130      140      150      160      170
gi-13621430. DETVYEVVIYVLYNEQSGALETNLVSNKLGETEKSELIFKQEYSEKTPPEPHQPDTEKEK
gi-19224134. VLMGGQSESVEFTKDTQTGMSGFSETVTIVEDTRPKLVFHFDDNNEPKVEEN-----REK
      610      620      630      640      650

      180      190      200      210
gi-13621430. POKKRNGLPSTGEMVSYVSALGIVLVATITLYSIYKCLKTSK
gi-19224134. PTKNITPILPATGDIENVLAFGLILSVLSIFSLLKKNKQNNKV
      660      670      680      690

```

! Distributed over 1 thread.
 ! Start time: Wed Sep 15 18:45:36 2004
 ! Completion time: Wed Sep 15 18:45:38 2004
 ! CPU time used:
 ! Database scan: 0:00:00.1
 ! Post-scan processing: 0:00:00.3
 ! Total CPU time: 0:00:00.4
 ! Output File: gi-13621430.fasta

FIGURE 59G

TYPE 3 pilus motifs
protein F2 like fibronectin-binding protein
 Length: 696-733

LPXTG

pilin motif consensus PK (X₇) K
E box consensus ETxAPxGY

SpyM3_0104/21909640	pilin motif	155 PKEKPIIYFK
	E box	398 YTFVETAAPDGY
SPs0106/28895018	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
SpyM18_0132/19745307	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY
orf84	pilin motif	269 PKEKPIIYFK
	E box	512 YTFVETAAPDGY

TYPE 4 pilus motifs
protein F2 like fibronectin-binding protein
 Length: 1161

LPXTG

pilin motif consensus PK (X₇₋₈) K
E box consensus YxLxETxAPxGY

The protein is longer than the proteins belonging to TYPE 3 and has 4 possible pilin motifs and 2 E boxes

19224141	pilin motifs	215 PKGISQDIPK
		571 PKGYQQVTEK
		156 PKMSVVSKEYGK
		674 PKYDAKNQEQYK
E boxes	563 YDLYETKAPKGY	
	940 YTFVETAAPDGY	

FIGURE 60

Formation of pili structures on GBS appears to be correlated to FACS values for surface expression of GBS80 protein

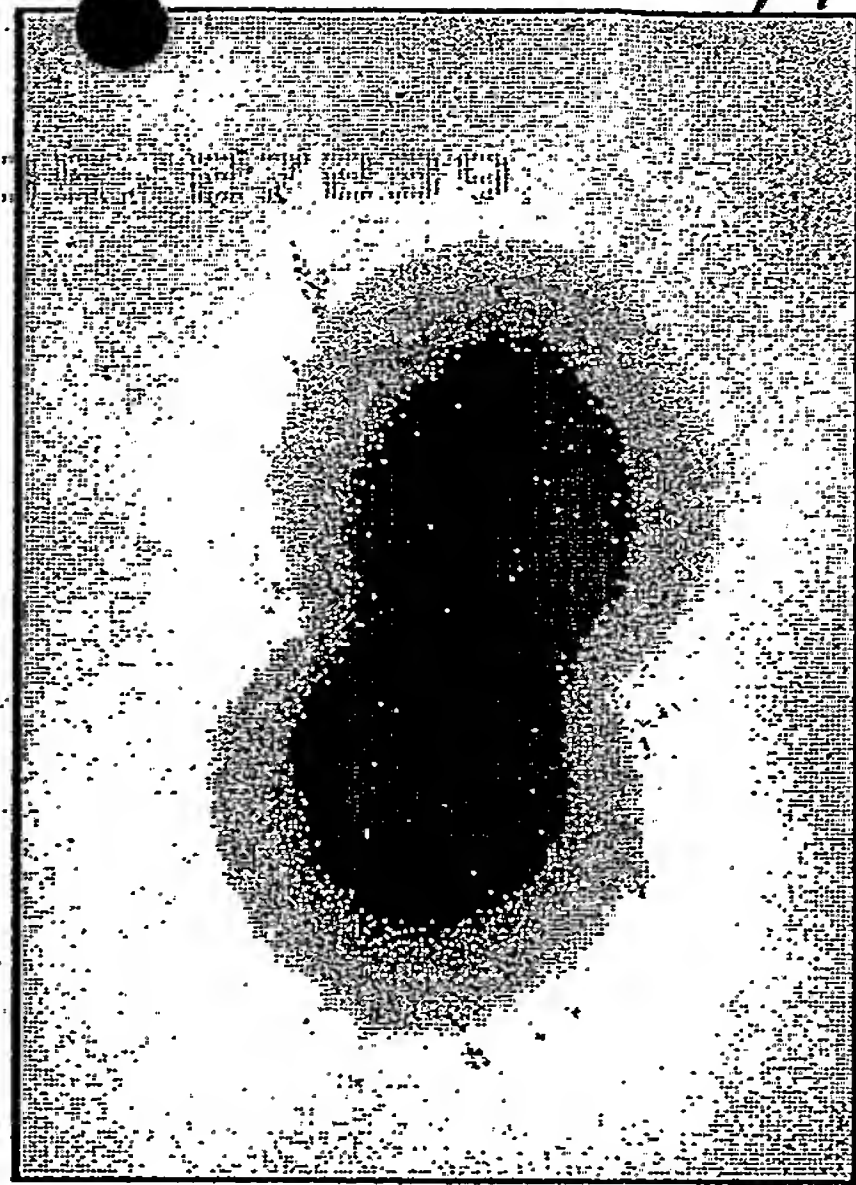
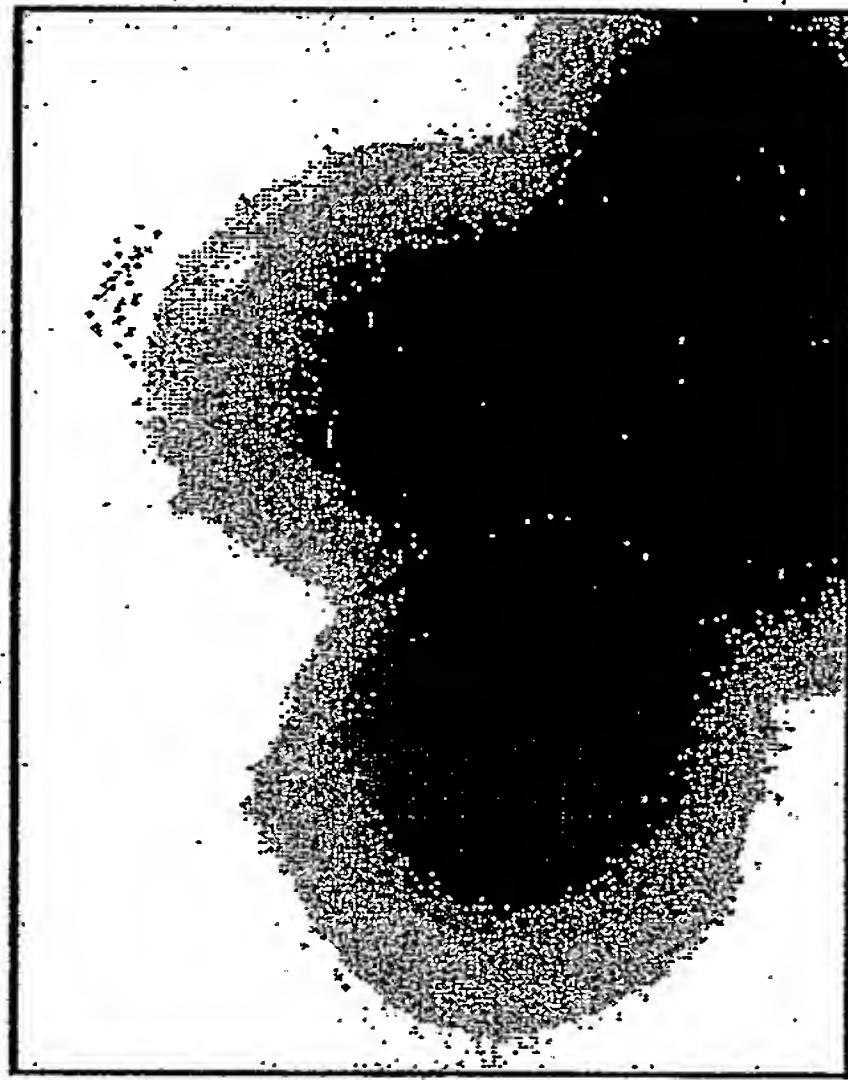
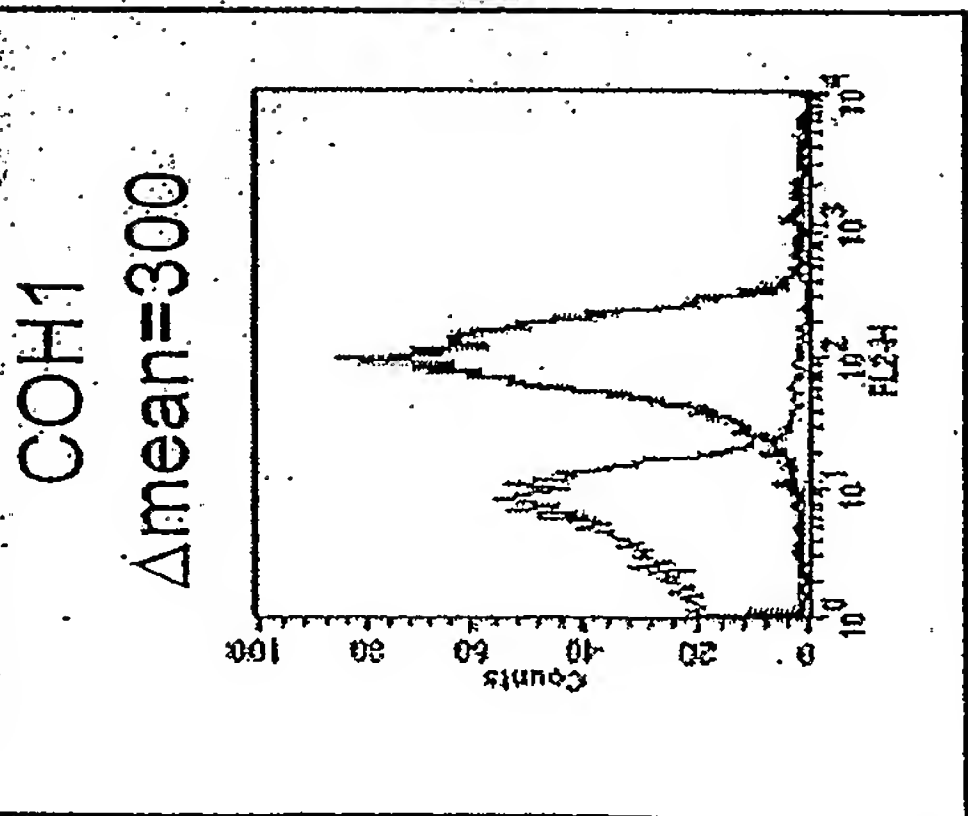
WO 2006/078318

PCT/US2005/027239

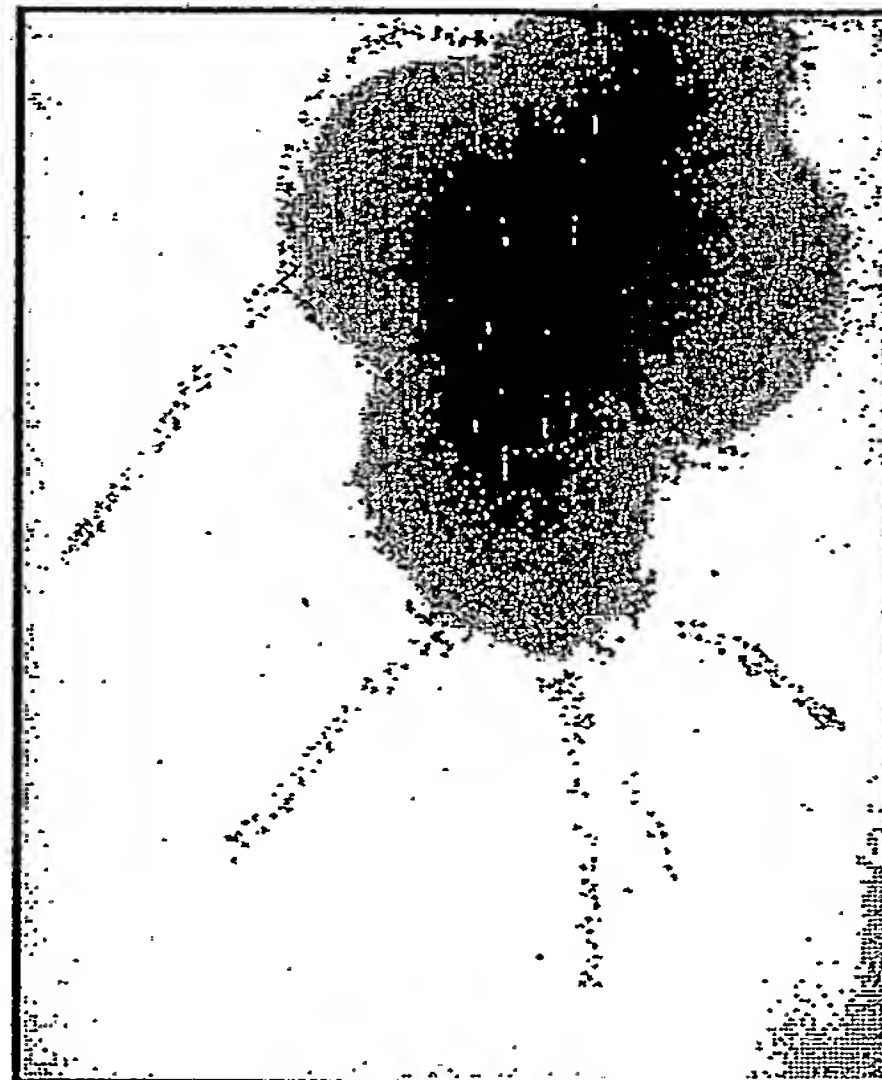
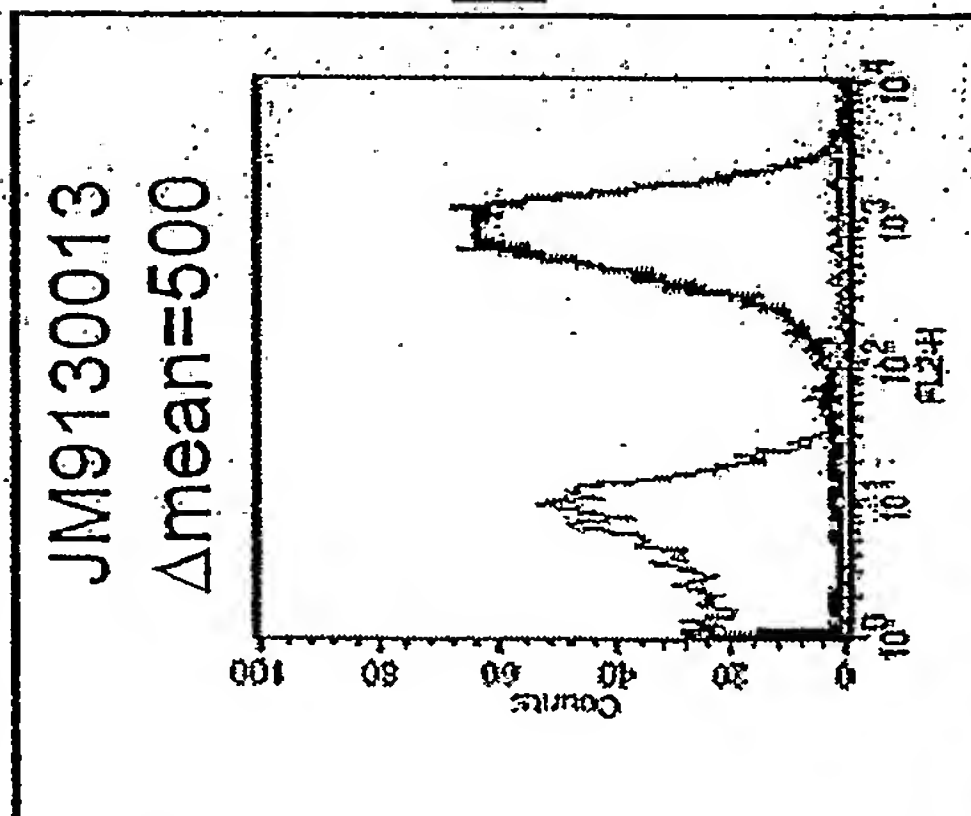
2/1/487

Immunogold Electron Microscopy

α -GBS80



α GBS80



PCT/US2005/027239

FIGURE 61

Surface exposure is capsule-dependent for GBS 322 but not for GBS 80

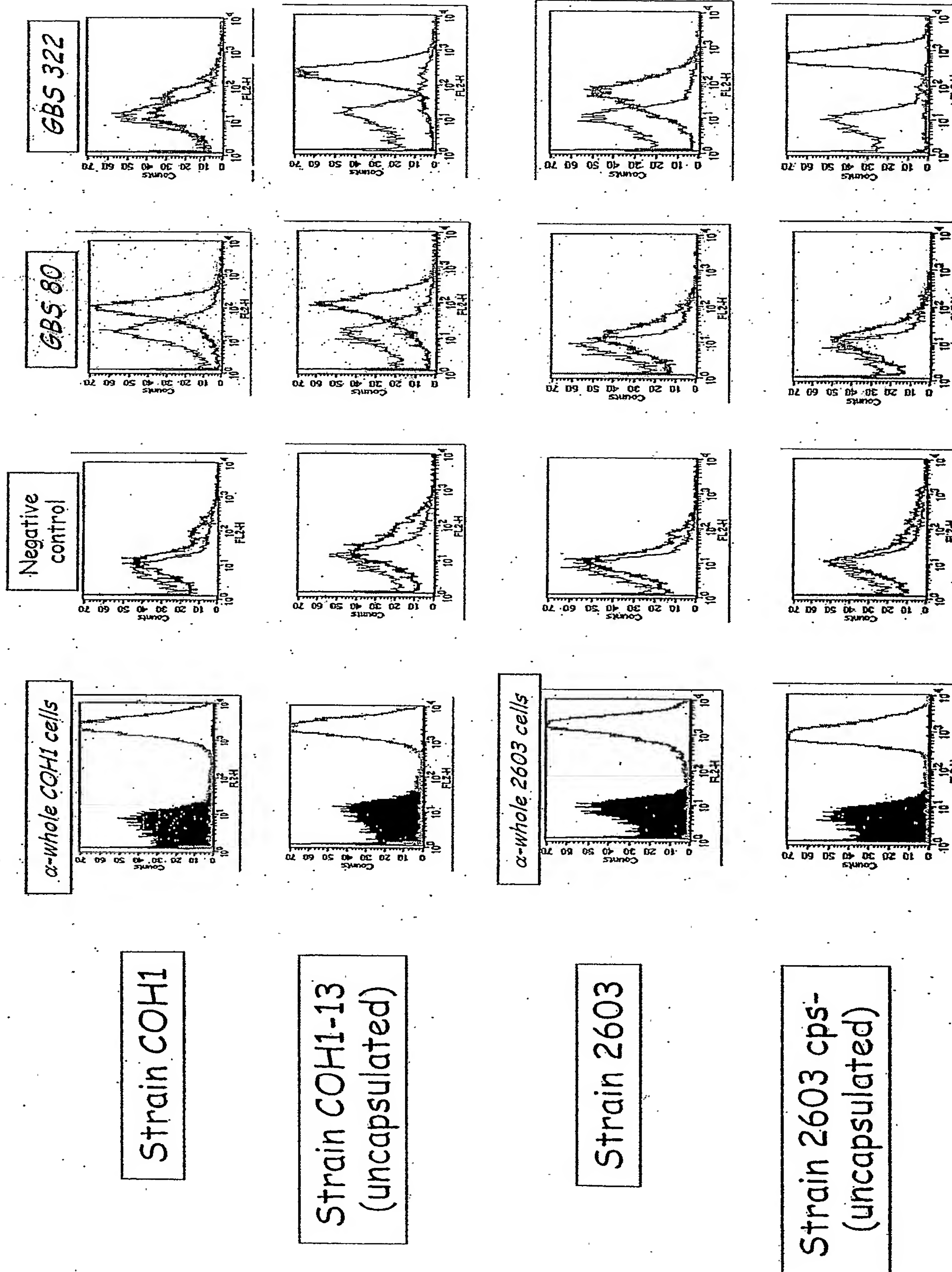
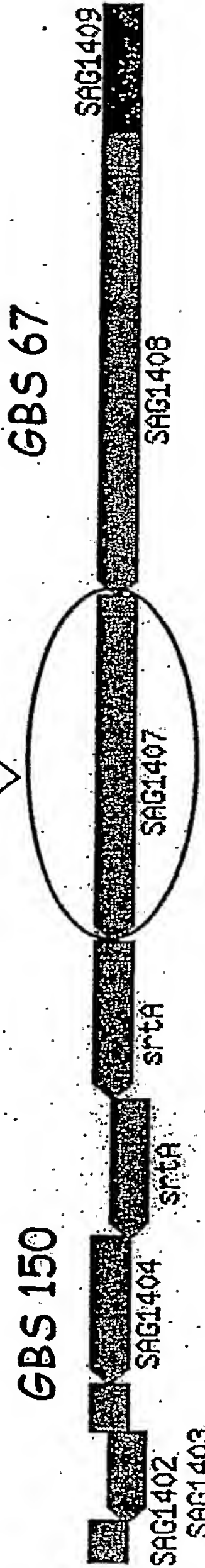


FIGURE 62

Adhesin island 2-
Operon gbs 67, 59, 150



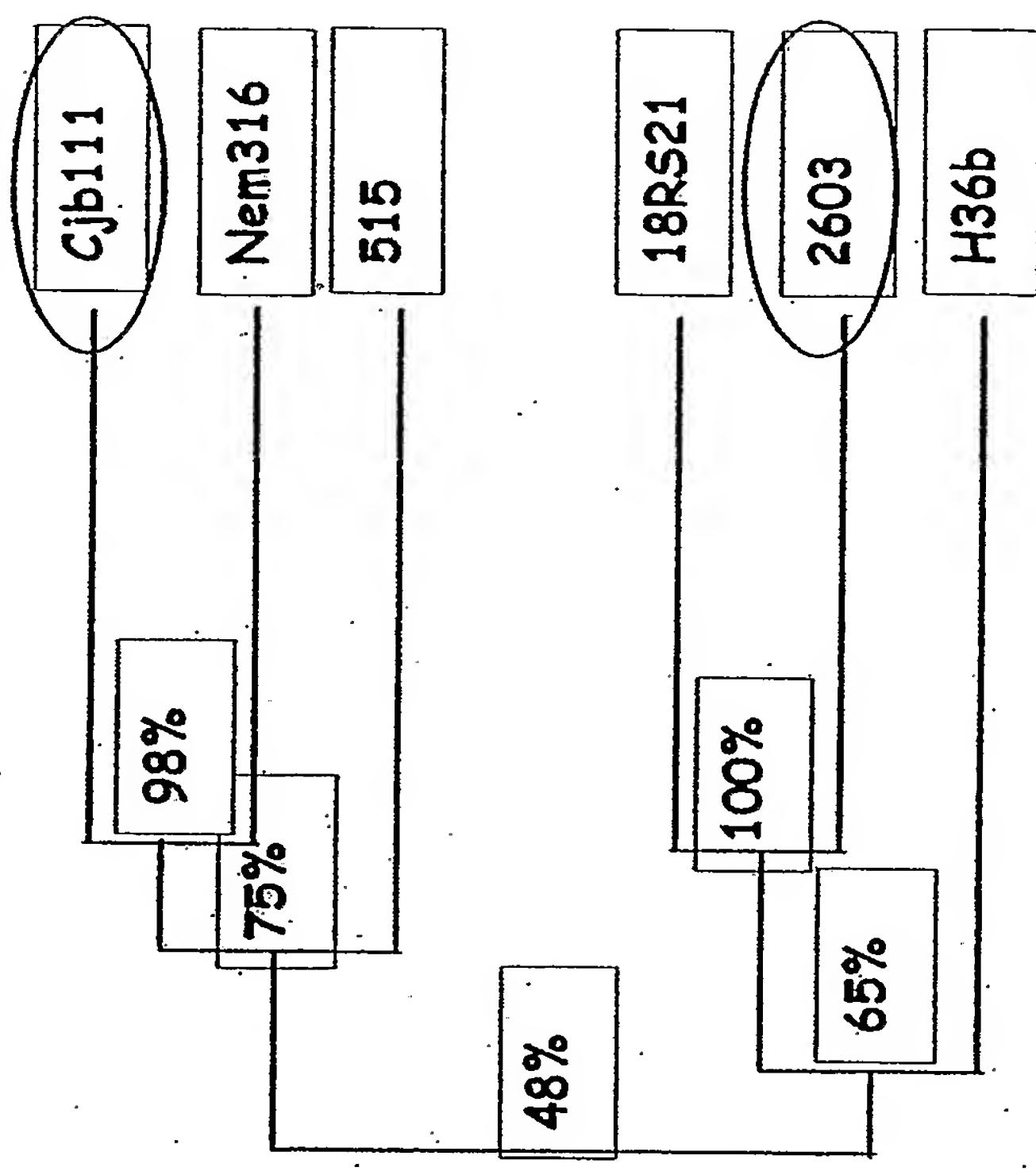
• 30 % identity with GBS 80

• By comparing GBS 59 amino acidic sequence of 2603 with that of other sequenced genomes, the following homologies are obtained:

2603	
nem316	47%
h36b	62%
515	48%
cjb111	48%
18rs21	100%
coh1	not present (Spb1)
A909	not present (Spb1)

• CGH: 1/20 GBS strains analyzed (18RS21)

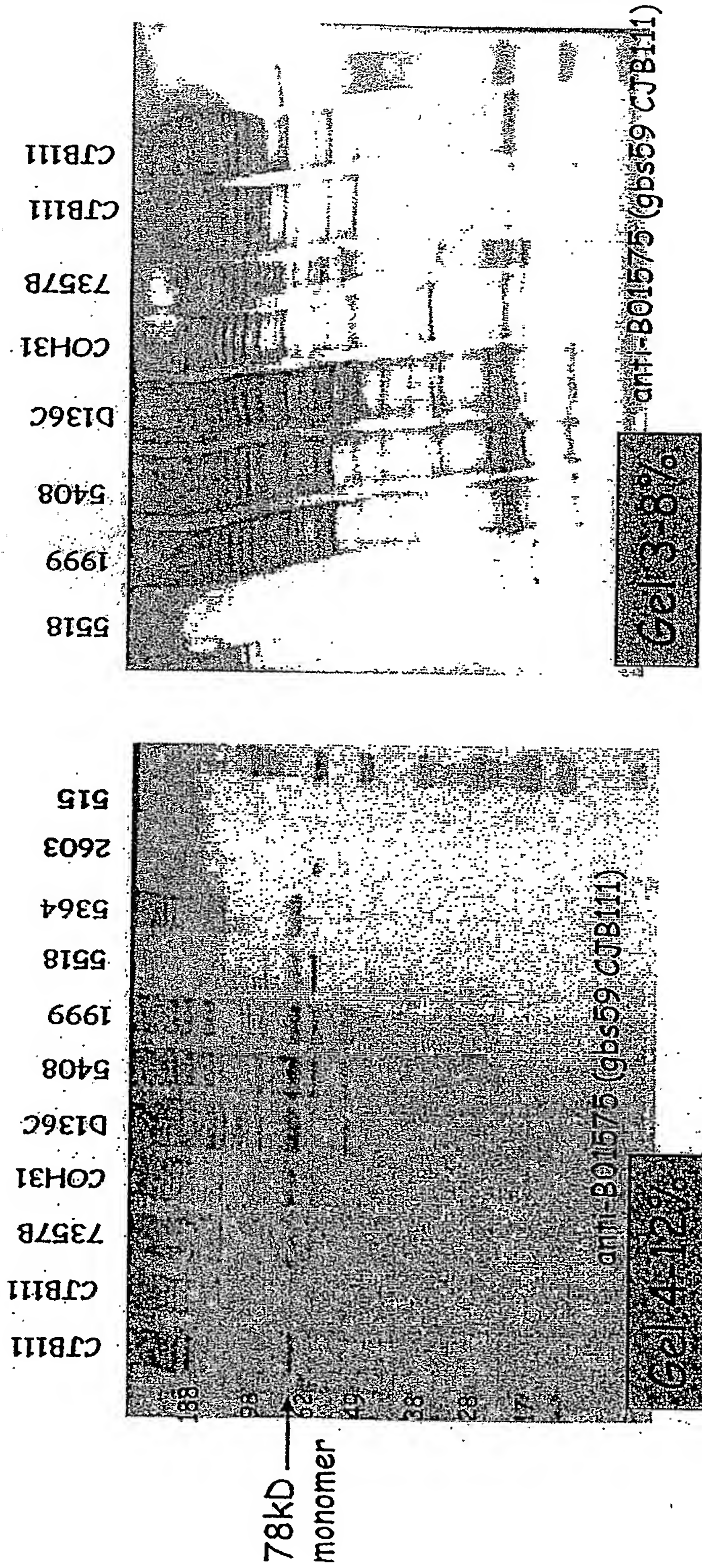
Two-by-two amino acidic sequence comparison



There seem to be two clusters suggesting the presence of two major isoforms

FIGURE 63

Western blotting with whole extracts derived from GBS strains



GBS 59 is part of a high molecular weight polymer (pilus)
 in GBS strains: CJB111, 7357b, coh31, d1363c, 5408, 1999,
 5364, 5518, 515

FIGURE 64

Western blotting with purified proteins and whole extracts derived from GBS strains

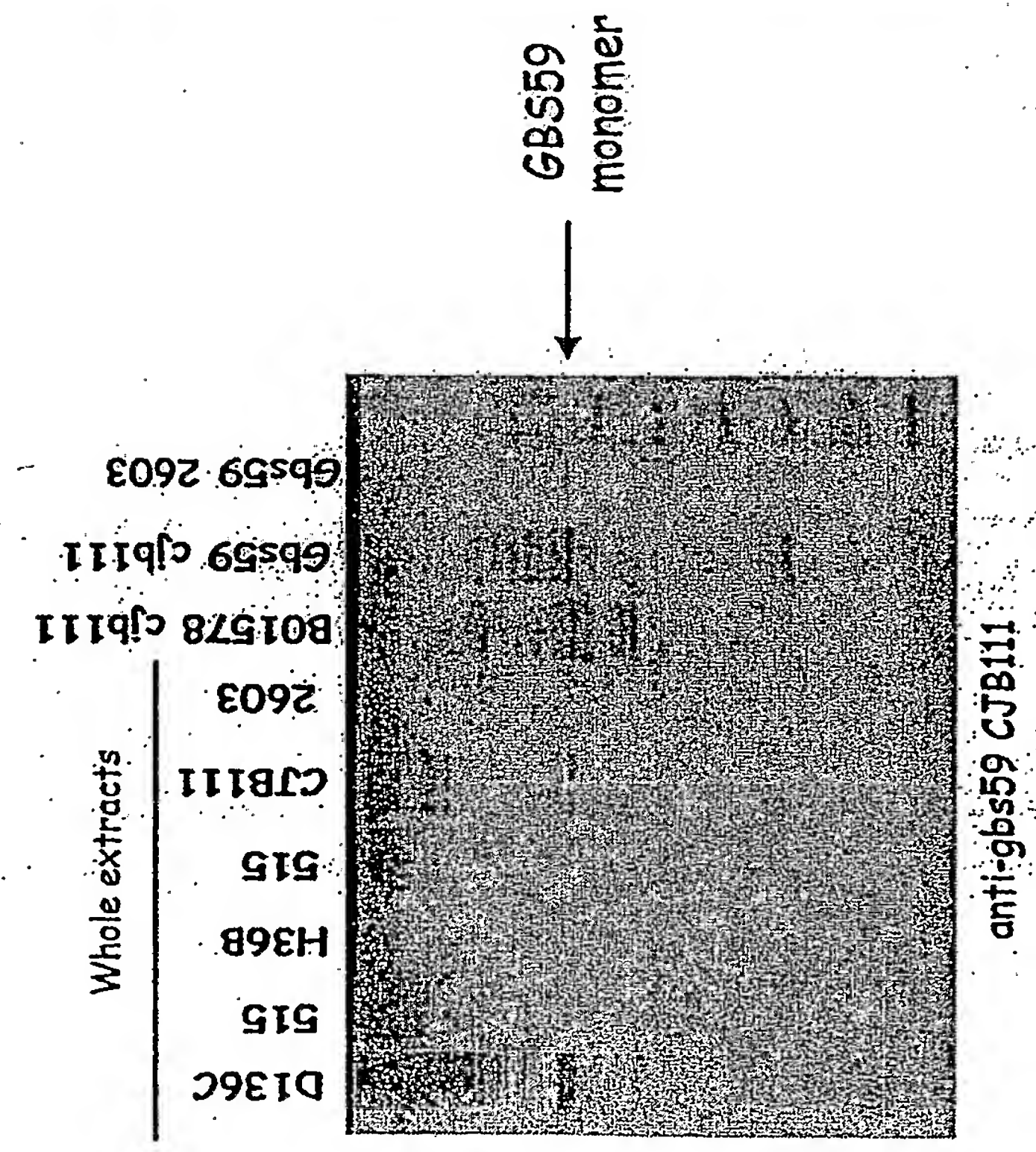
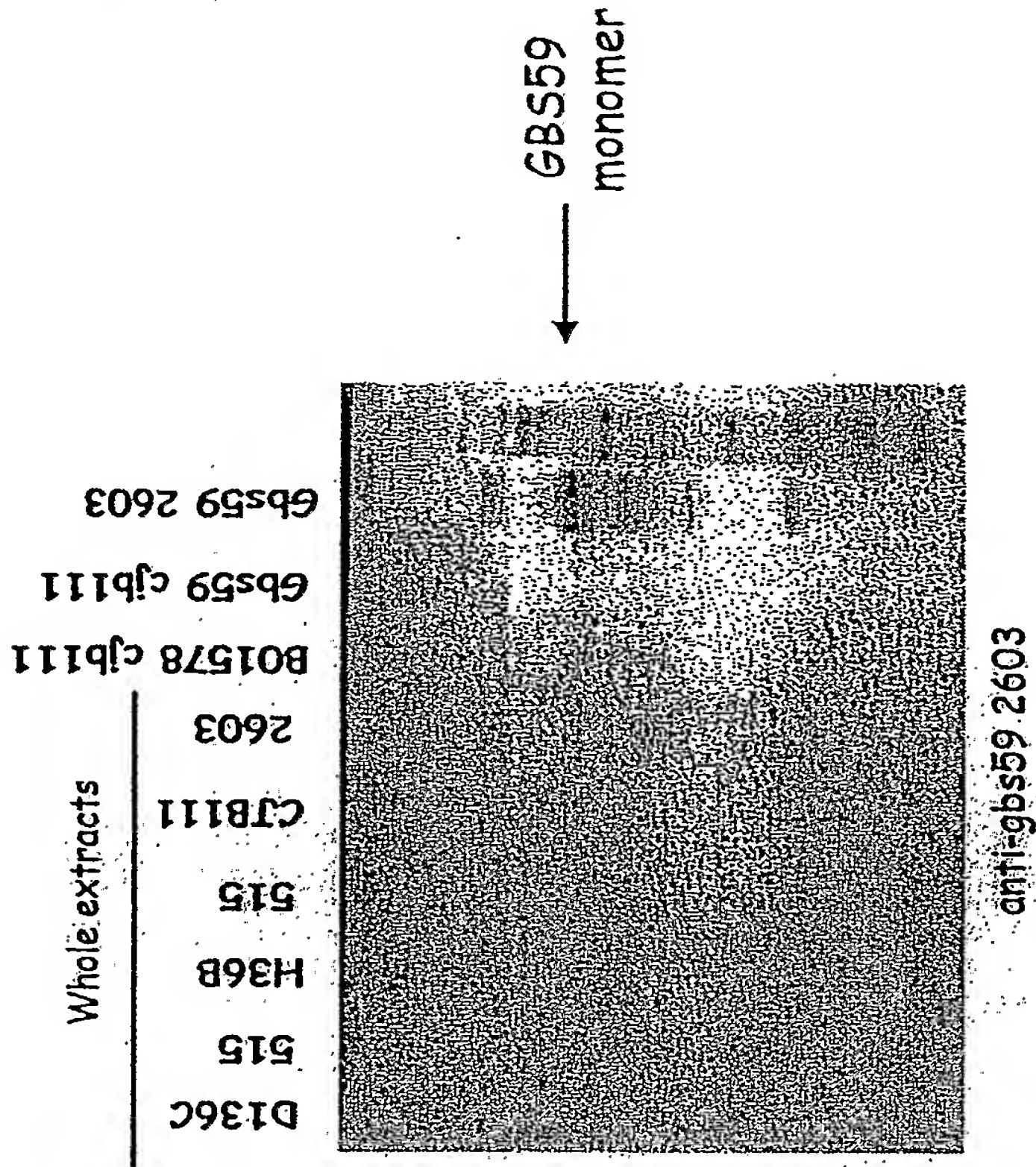
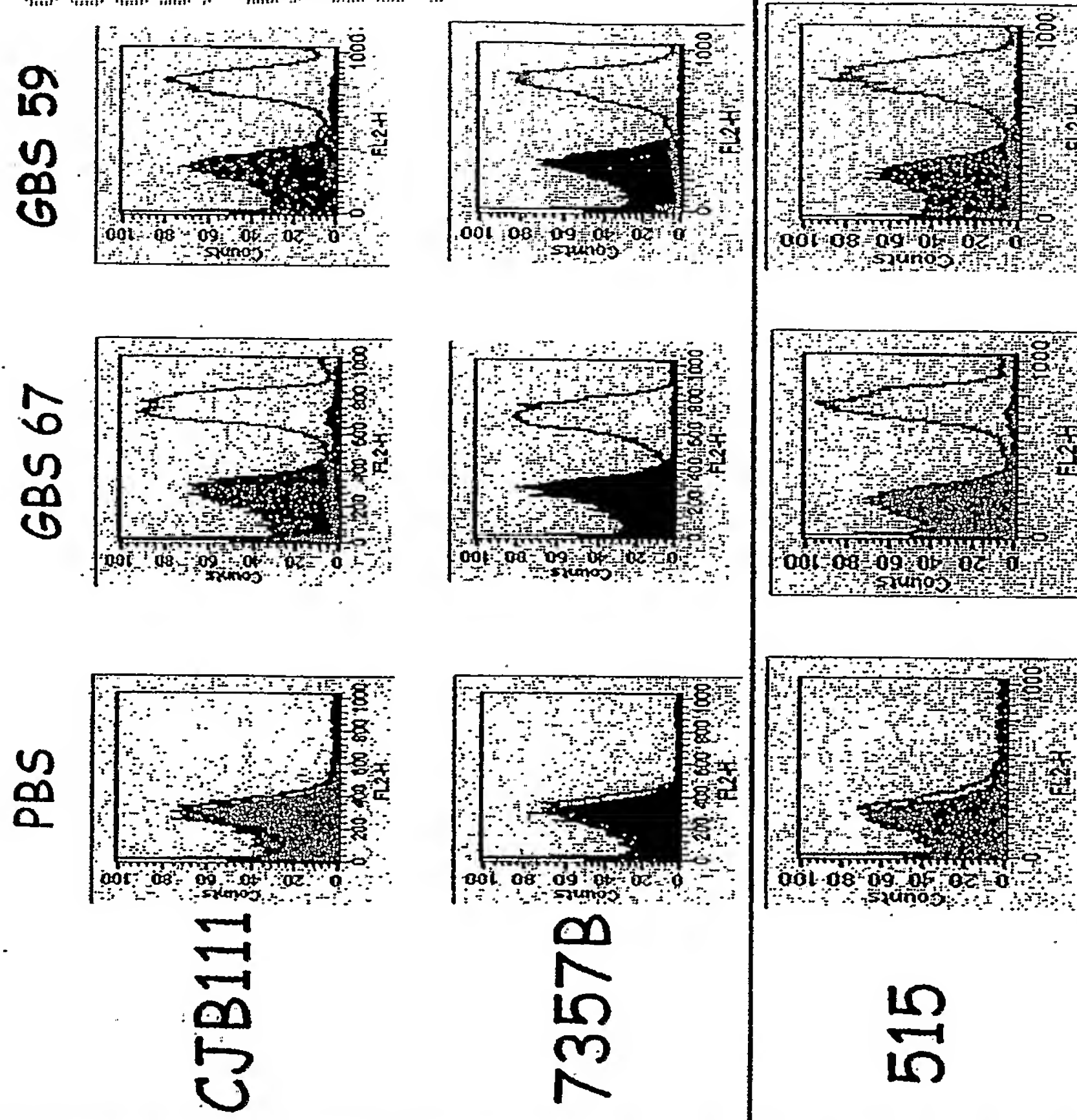


FIGURE 65

FACS analysis using mouse antiserum after immunization with BO1575 (gbs59) from CJB111 genome

GBS strains	Type	FACS (Δ Mean)	GBS 59
DK1		565	
DK8		559	
Davis	Ia	577	
515		583	
2986		443	
5551		524	
7357b-		596	
5518		190	
D136C	III	504	
COH31		505	
DK21	II	249	
CJB111		493	
5364	V	593	
2110		590	
1999		594	
2210		636	
5408		537	
1169		227	



Where present GBS 59 is a highly exposed on the GBS surface

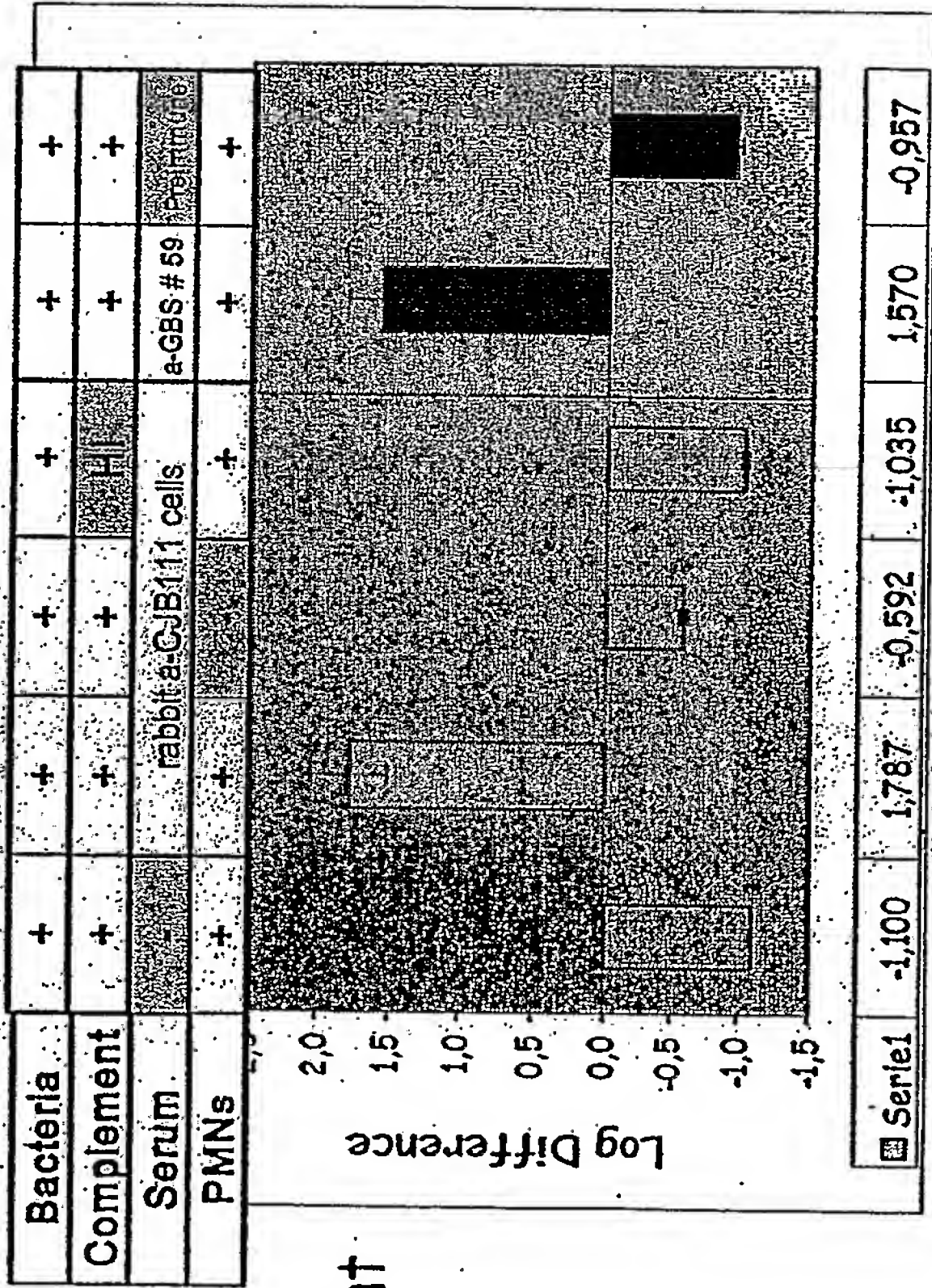
FIGURE 66

Opsonophagocytosis assays: B01575 (gbs59-cjb111)

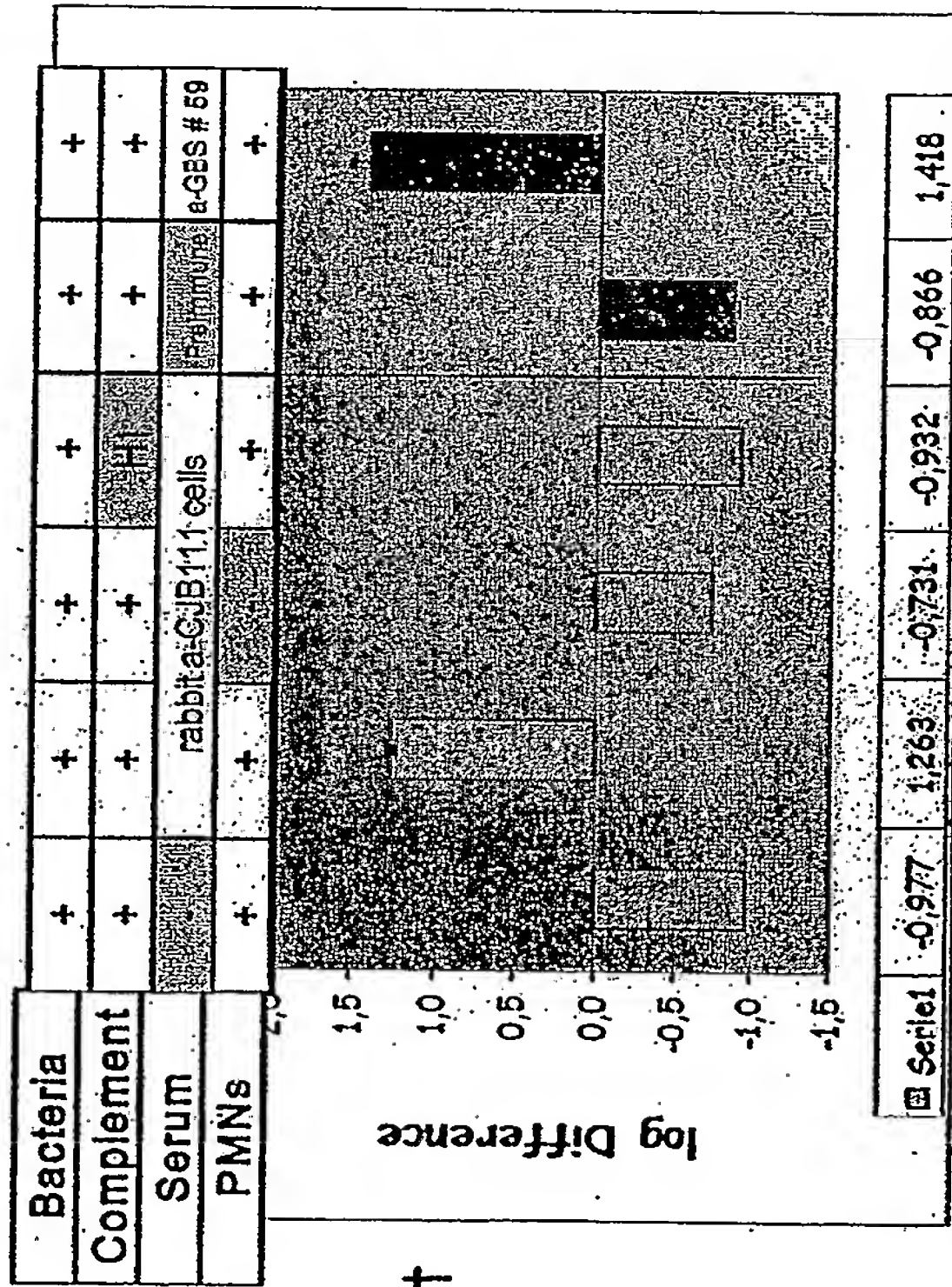
PCT/US05/27239

- CJB111 GBS strain type Ia
- Baby rabbit complement
- Human PMNs
- Positive control: anti-type V cells (rabbit serum anti fixed type V cells)

Antibodies against B01575 (cjb111) are opsonic for cjb111 GBS strain serotype V



I experiment



II experiment

FIGURE 67

Association GBS 80-104 WB α -80, α -104 JM9130013 Total Extract

Mab α -80 A477
 Mab α -80 19G4178
 Mab α -104 15H13149
 Mab α -104 12A7167
 Mab α -104 H2132
 Mab α -104 14F3173
 α -104 POLIC.
 α -80 POLIC.
 Mab α -80 19F177

FACS	(Δ MEAN)
GBS 80	597
GBS 104	446

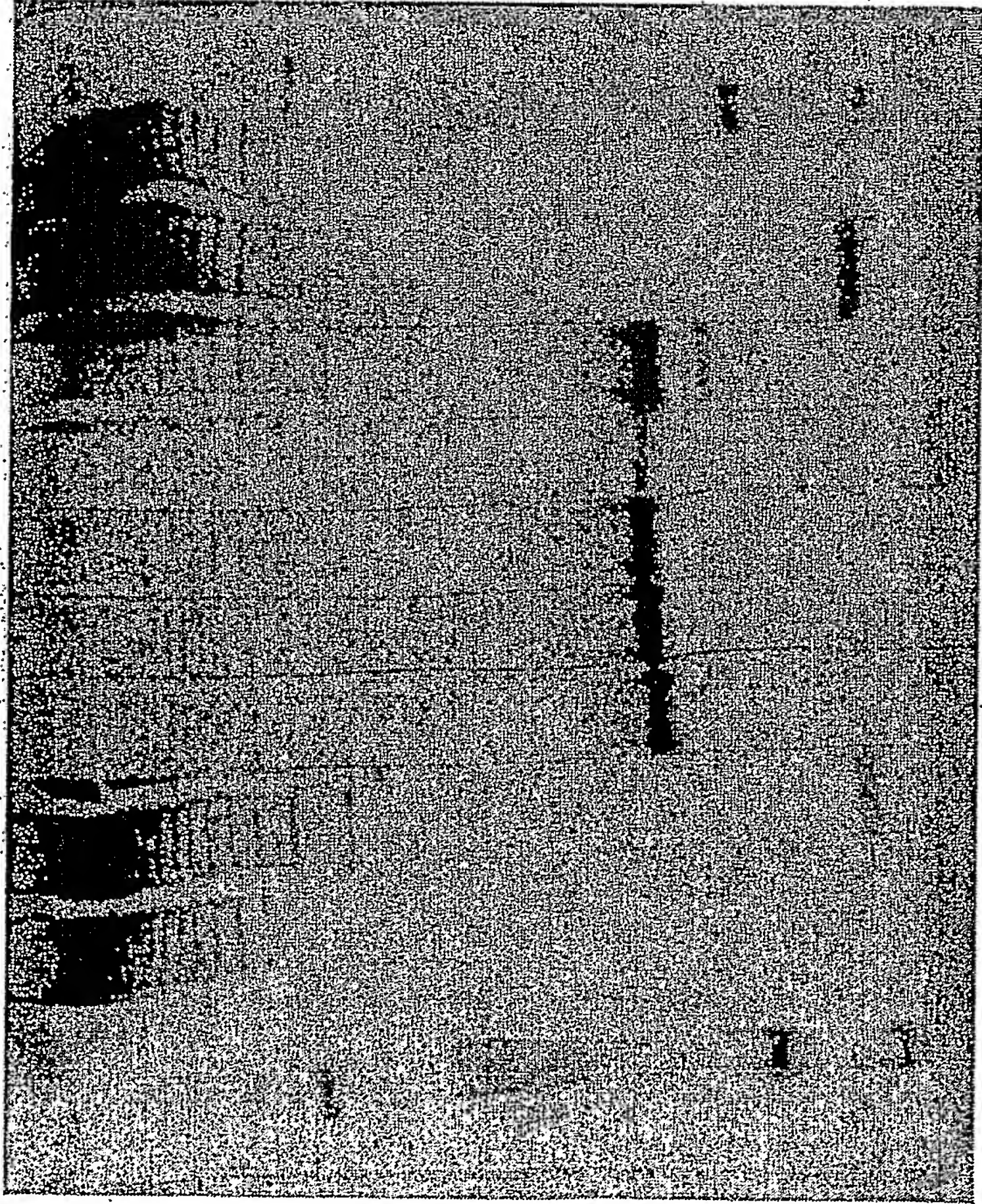
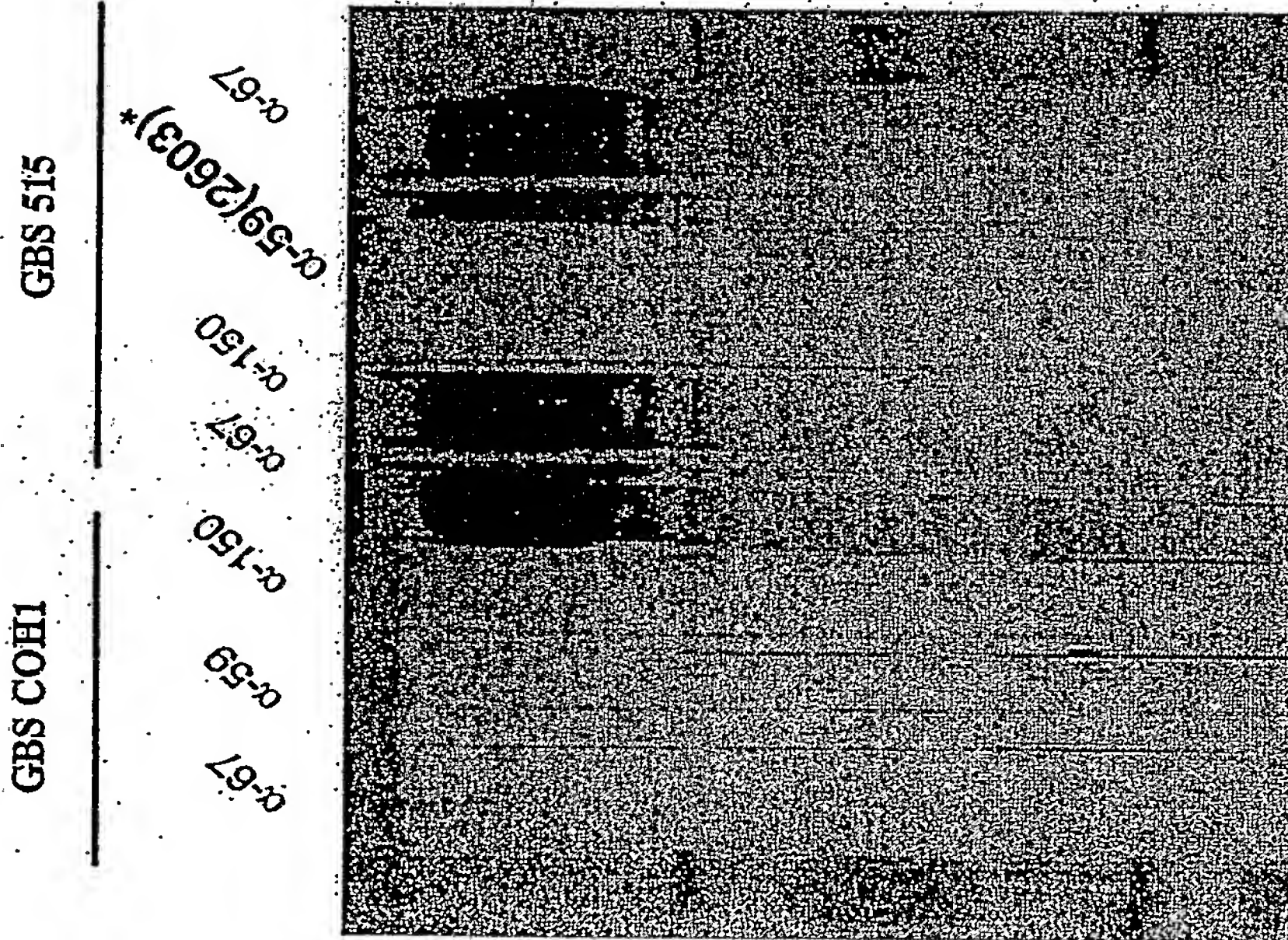


FIGURE 68

WB GBS 515 Total Extract

α -67; α -150; α -59 (2603)



Controls:
 GBS COH1 total extracts

• anti-gbs59 mouse serum after immunization with SAG1407 (GBS 59) from 2603 genome

GBS 67 and GBS 150 are parts of a high molecular weight polymer (pilus) in 515 GBS strain

FIGURE 69

Western Blotting ko GBS67
from 515 genome (clone 1.45)

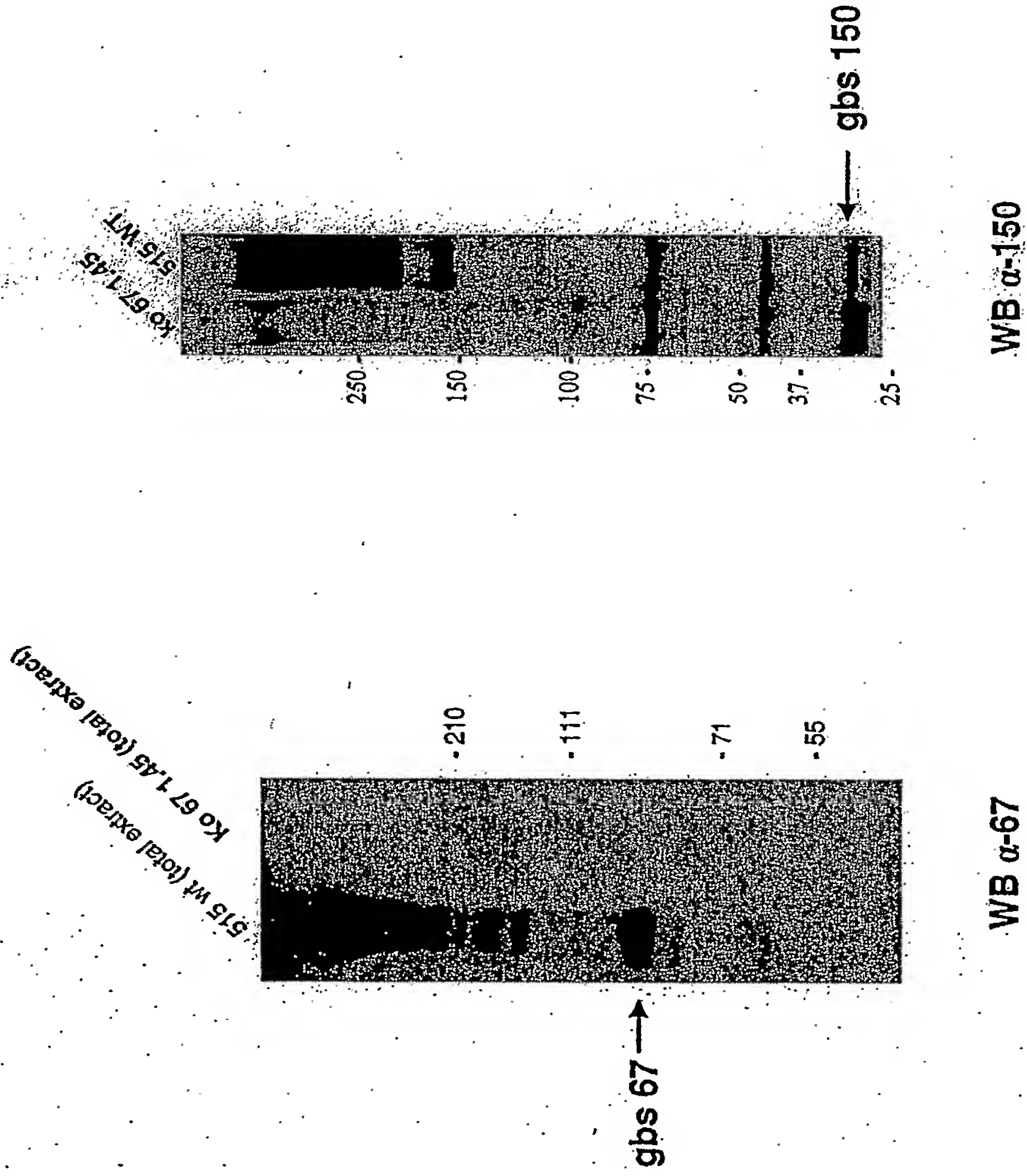


FIGURE 70

FACS GBS 515 Δ67

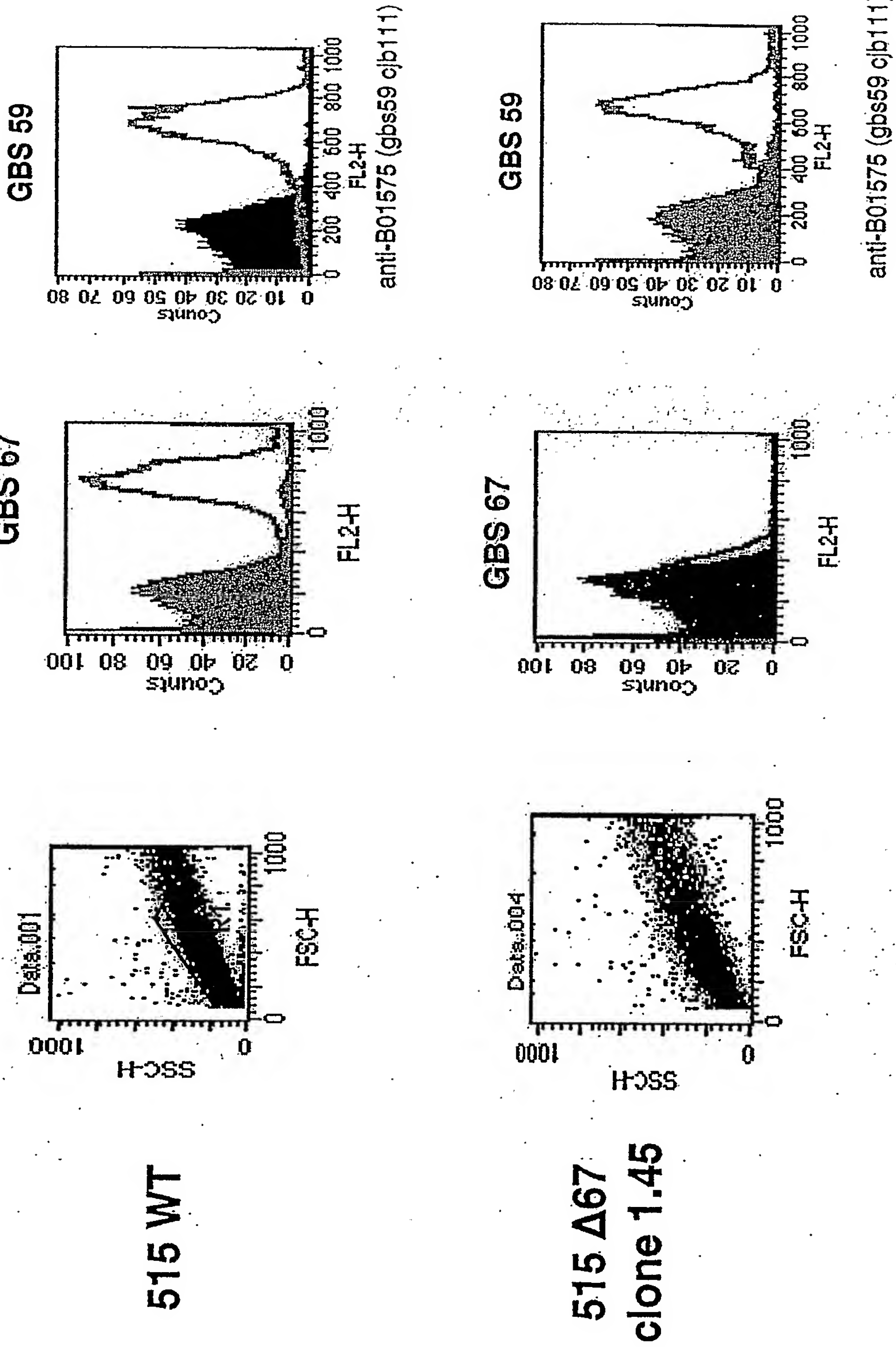
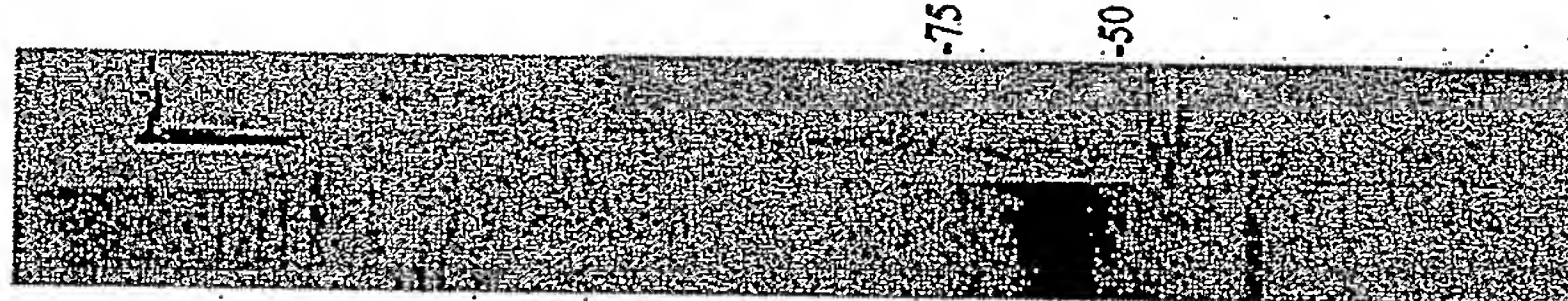


FIGURE 71

Complementation of GBS 515 KO 67 with pAM401-gbs80

KO 67 1.45
515 WT



GBS 80 forms a high molecular weight complex (pilus) in absence of GBS 67

WB α -80

FIGURE 72

spyM6_0159 type 1 pilus present in M6

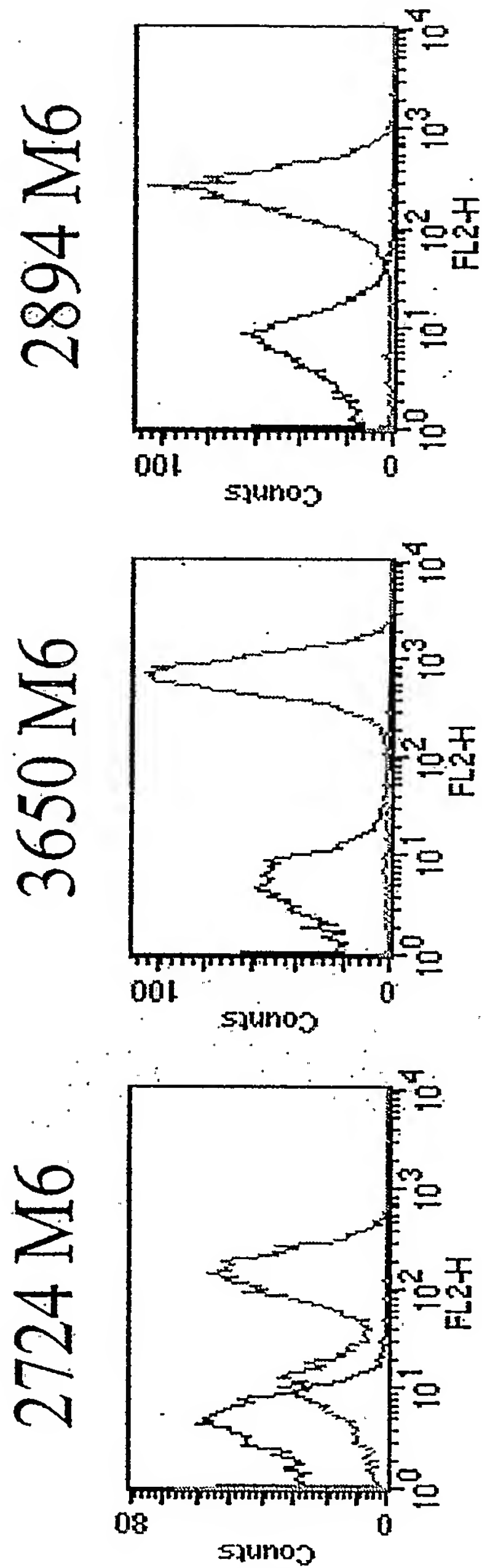


Figure 73

PCT/US05/27239 224/487

spyM6_0160 type 1 pilus present in M6

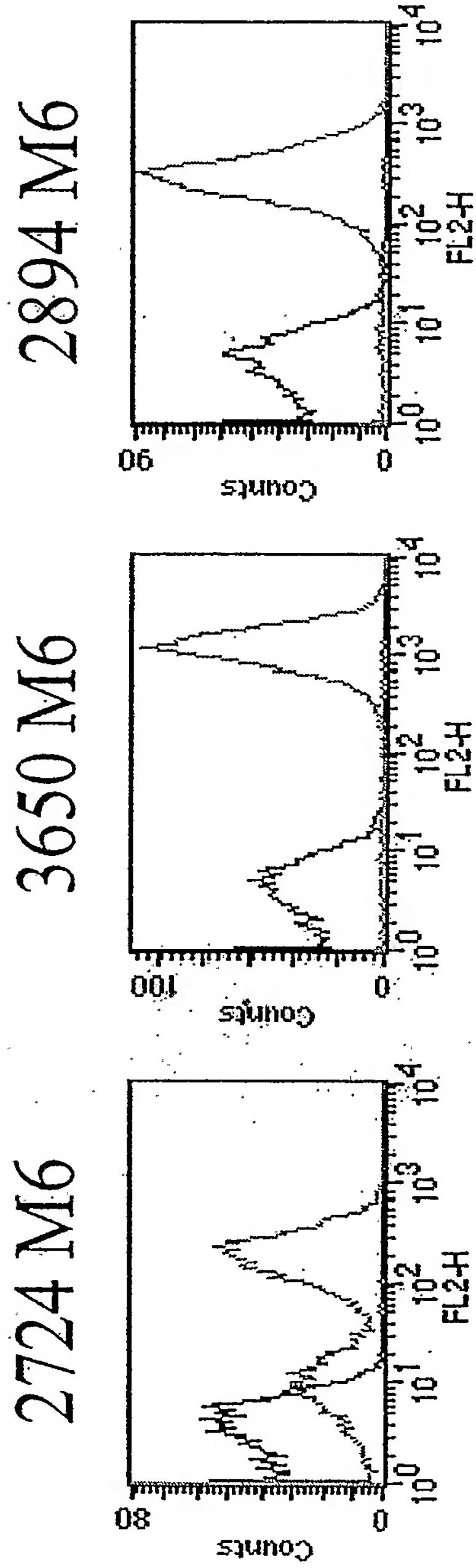


Figure 74

225/487

Gas15 type 2 pilus present in M1

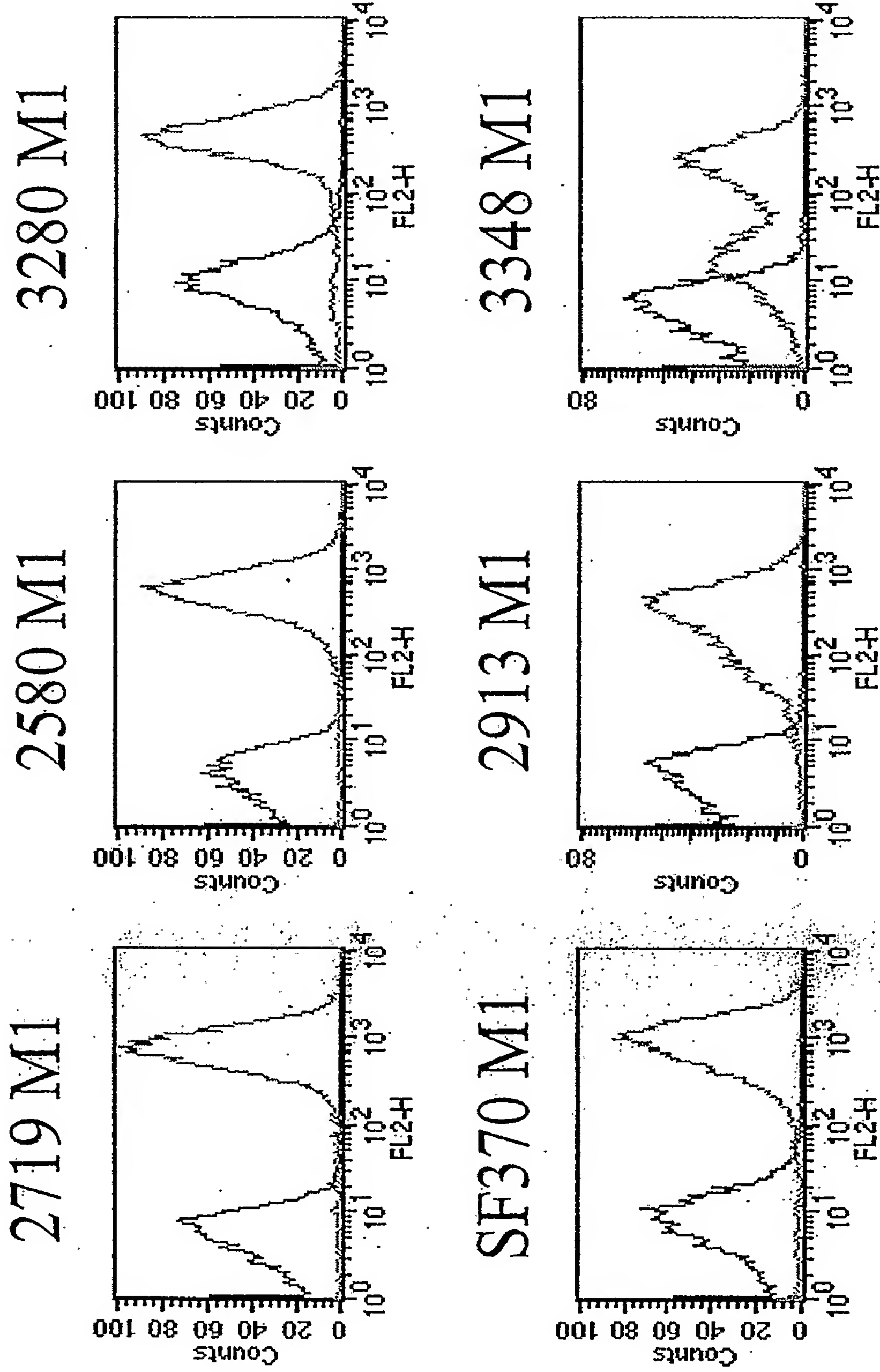


Figure 75

Gas16 type 2 pilus present in M1

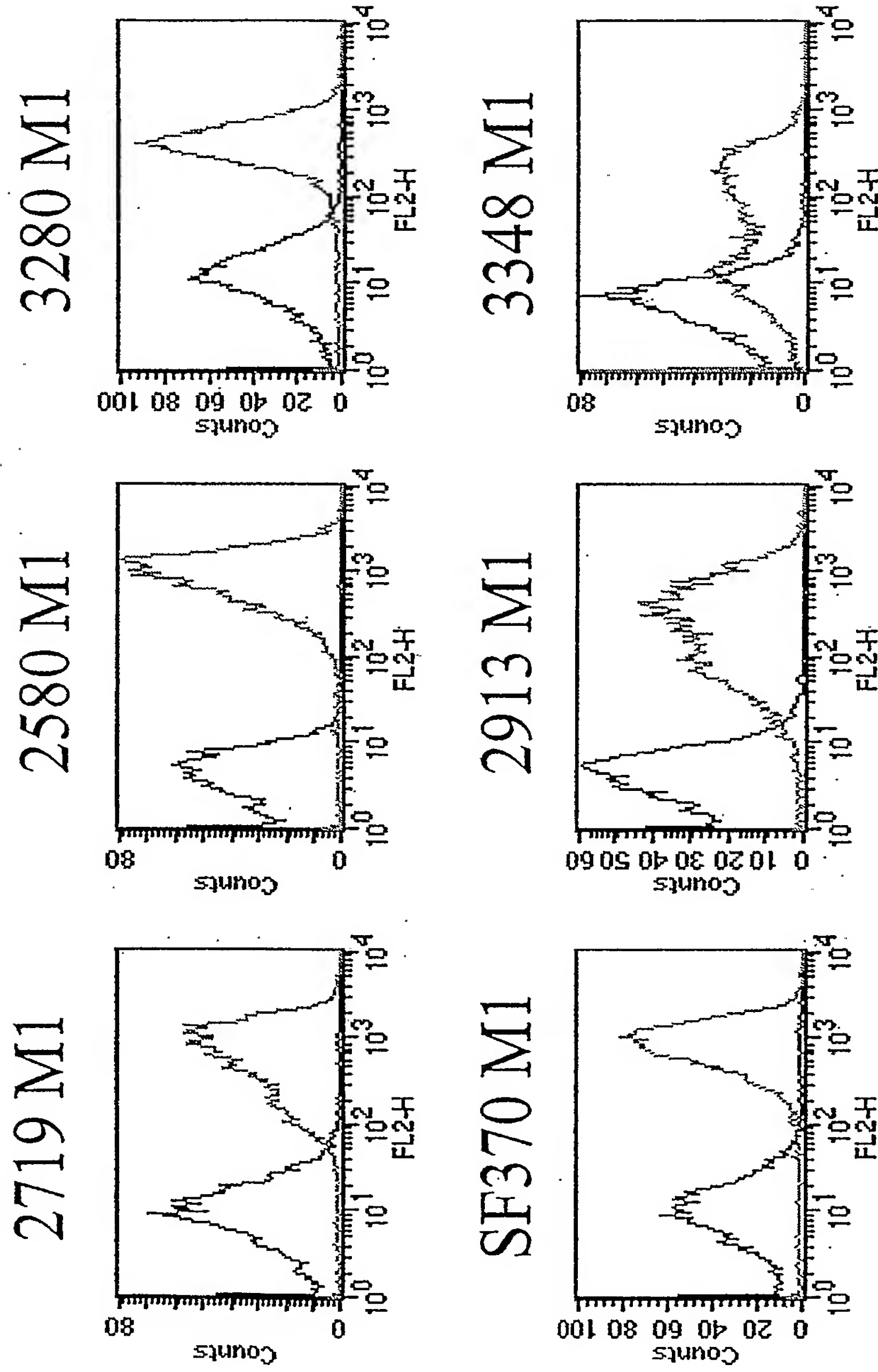


Figure 76

Gas18 serum 1 type 2 pilus present in M1

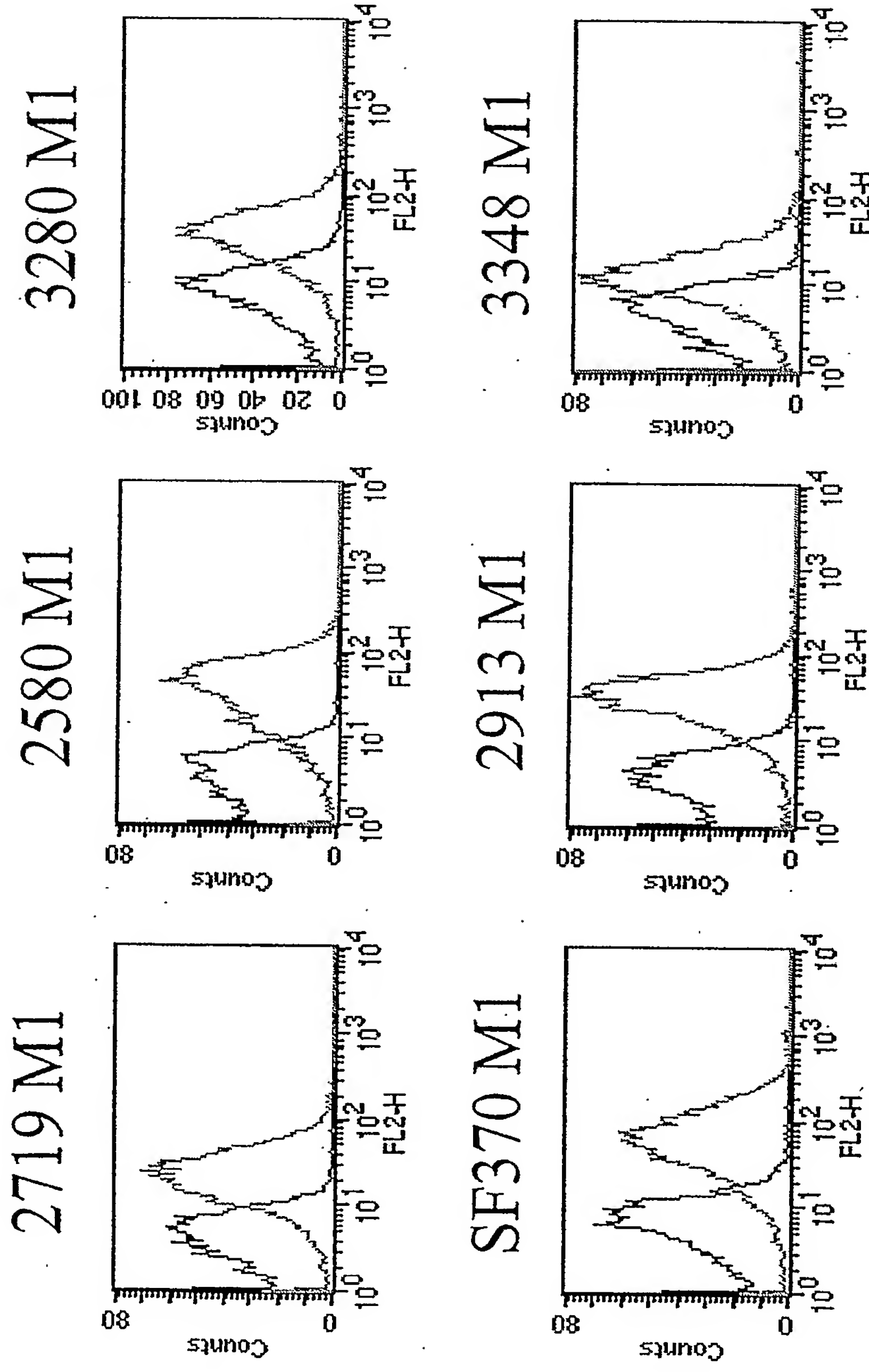


Figure 77

Gas18 serum 2 type 2 pilus present in M1

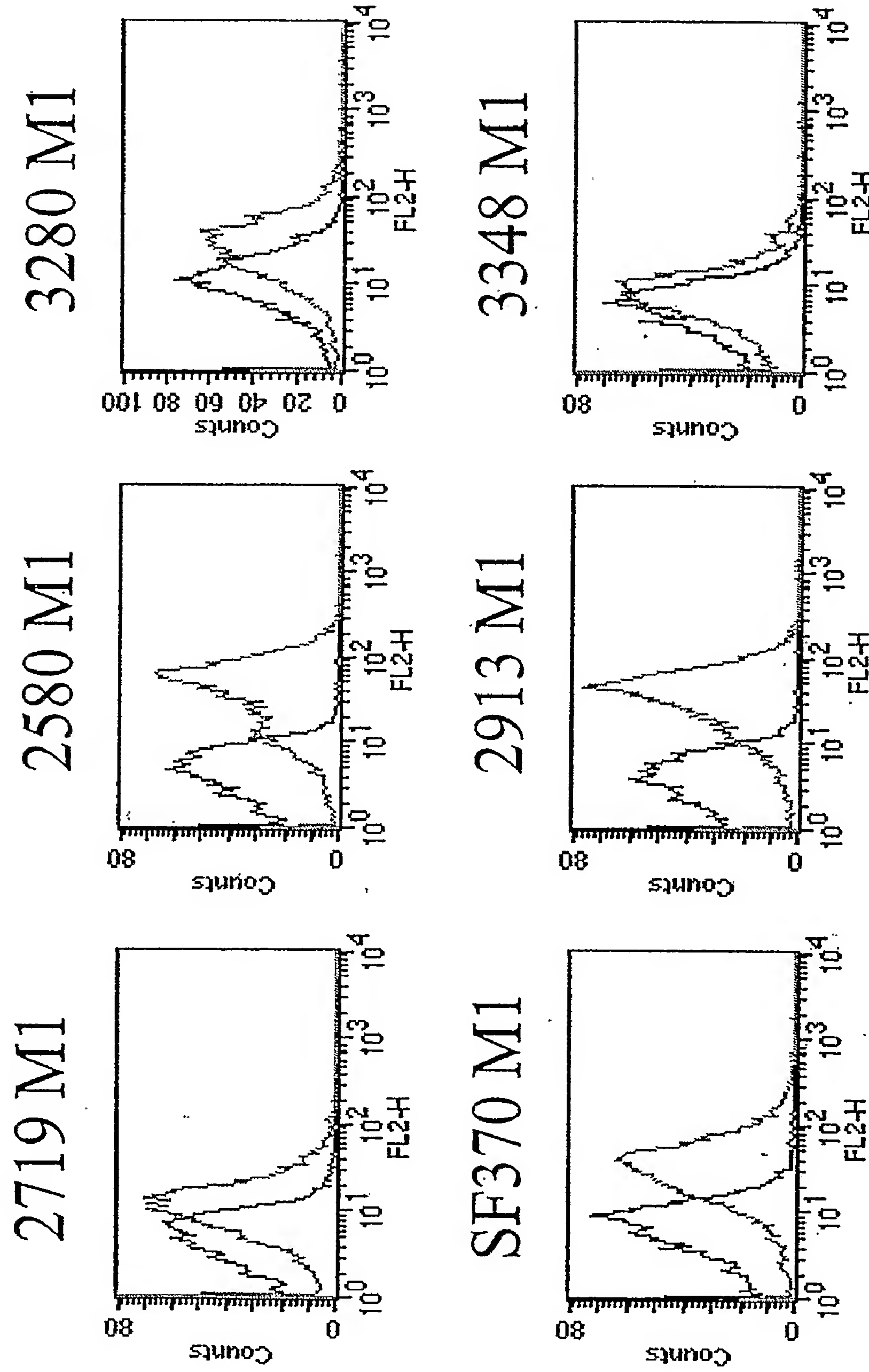


Figure 78

229/487

Gas16p2 type 2 pilus present in M1

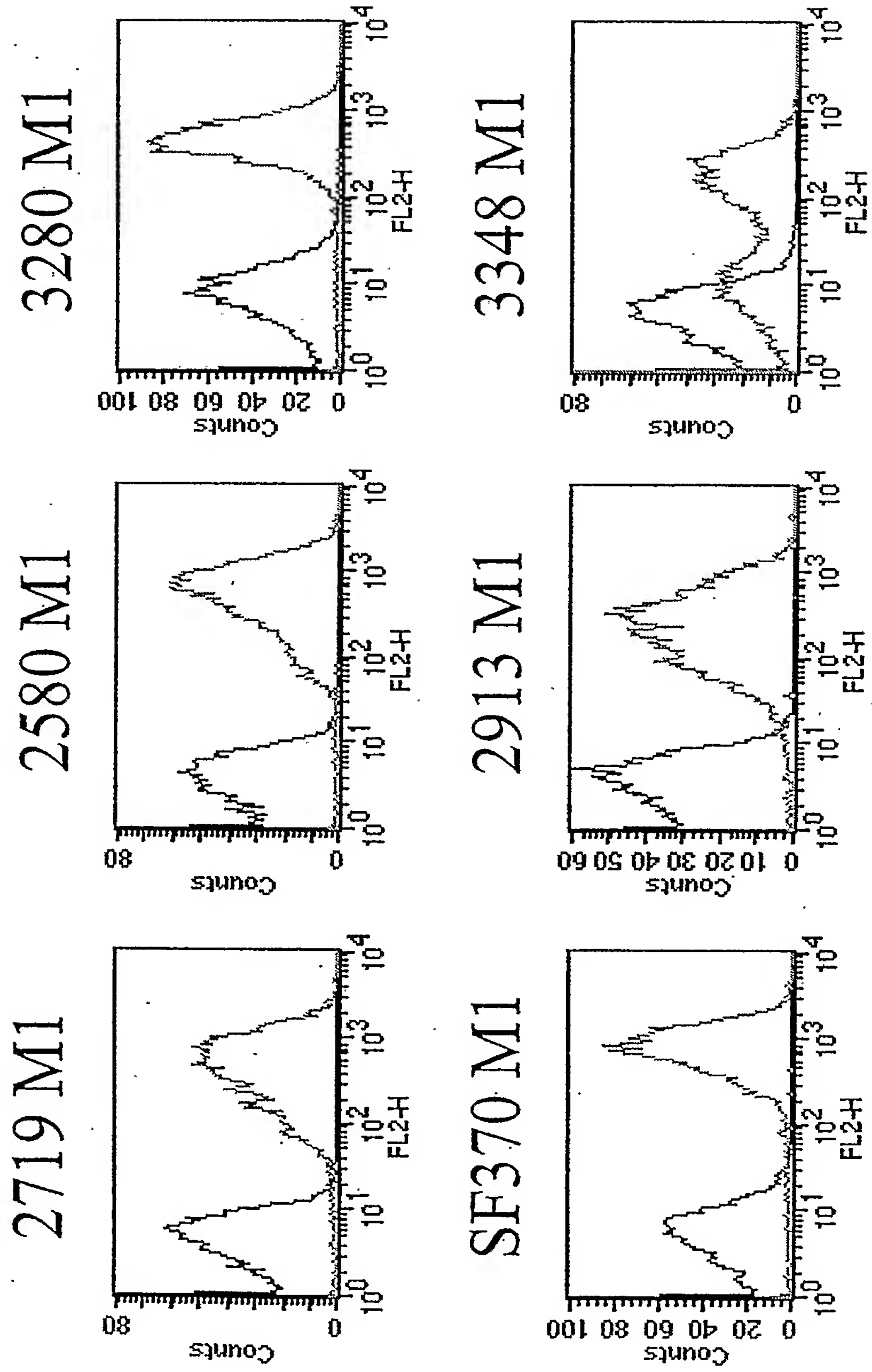


Figure 79

230/487

spyM3_0098 type 3 pilus present in M3

2721 M3

3135 M3

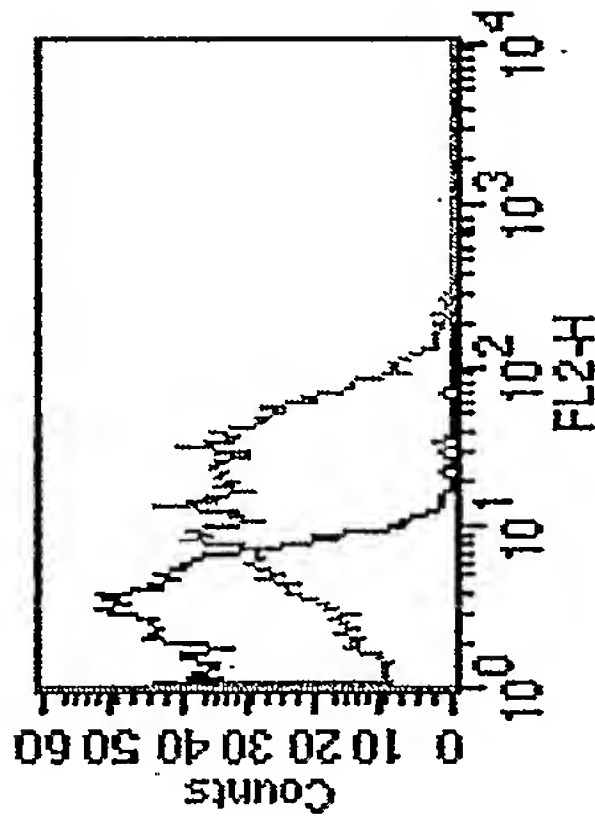
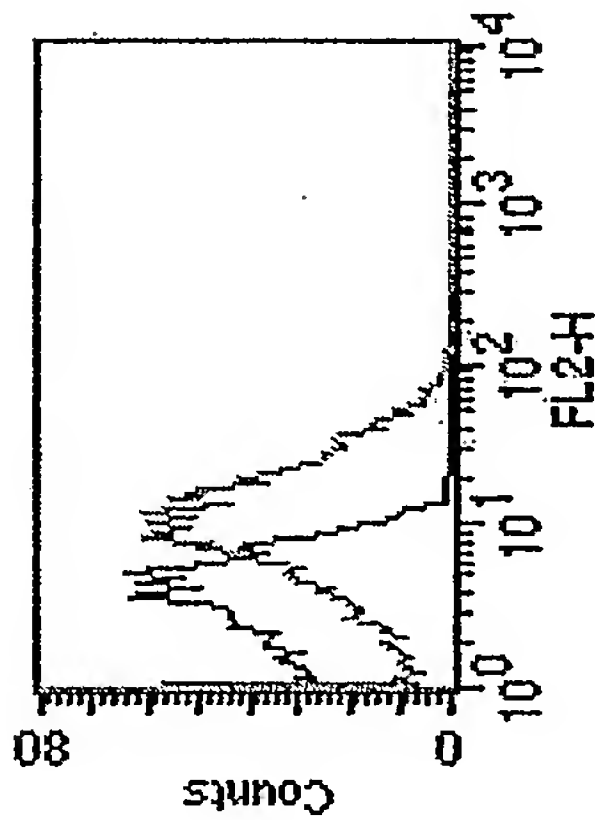
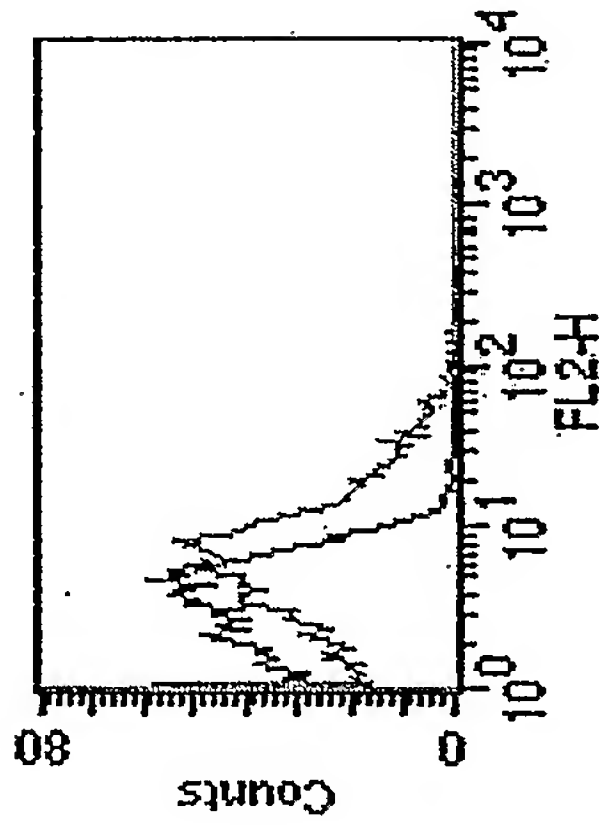


Figure 80

231/487

spyM3_0100 type 3 pilus present in M3

2721 M3



3135 M3

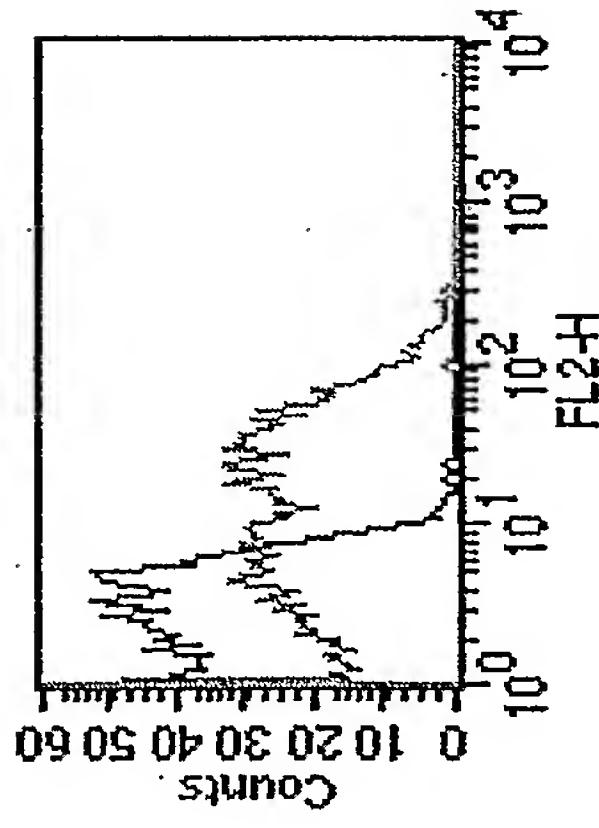
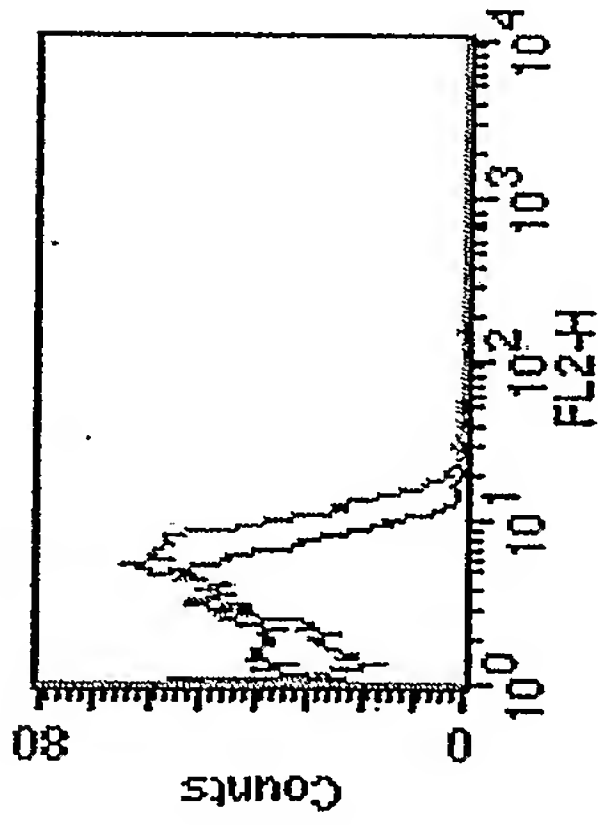


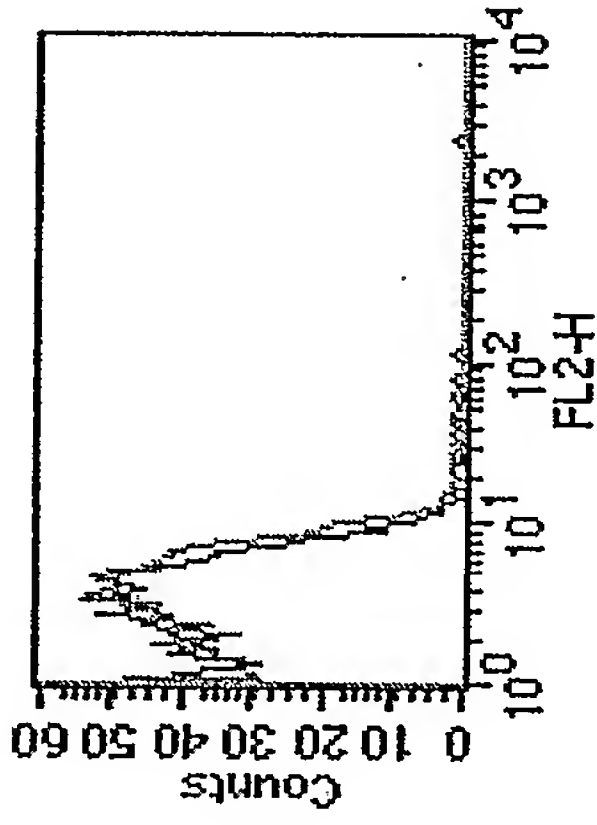
Figure 81

spyM3_0102 type 3 pilus present in M3

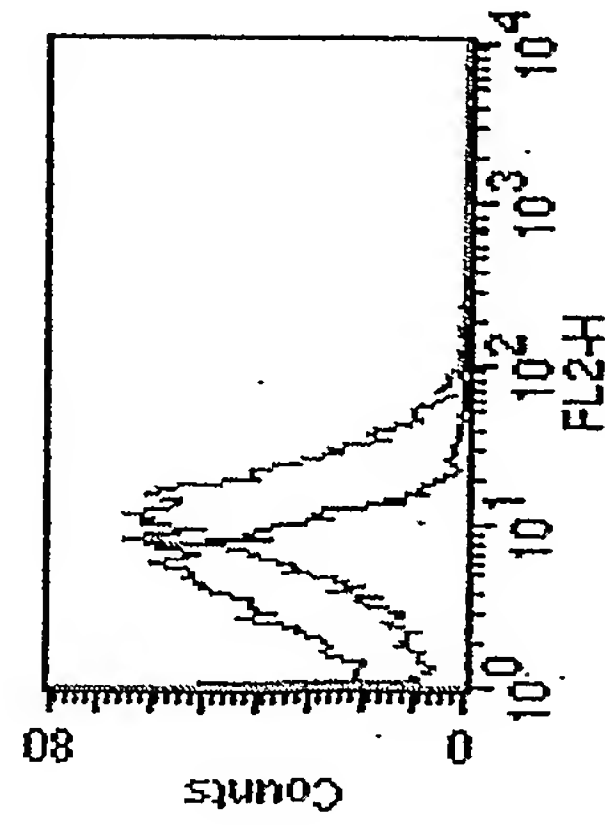
2721 M3



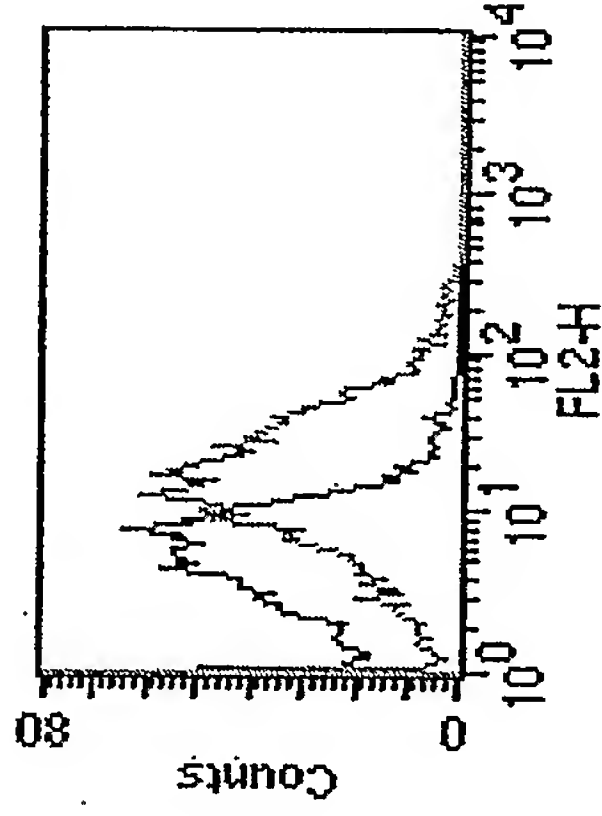
3135 M3



2724 M6



3650 M6



2894 M6

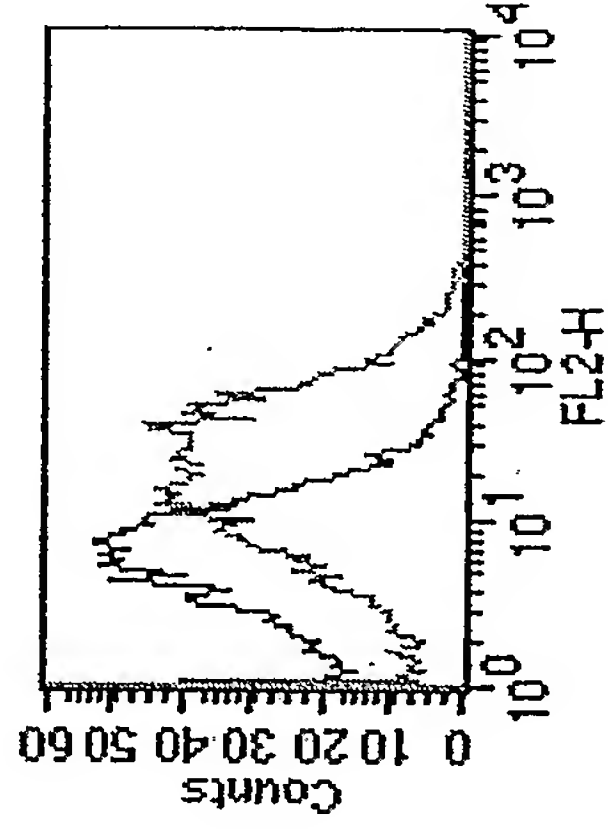
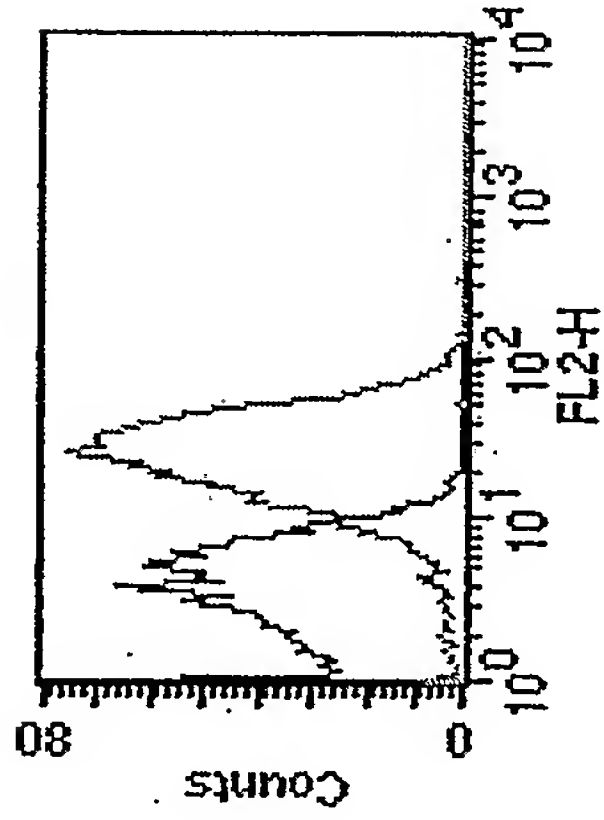


Figure 82

spyM3_0104 type 3 pilus present in M3

2721 M3



3135 M3

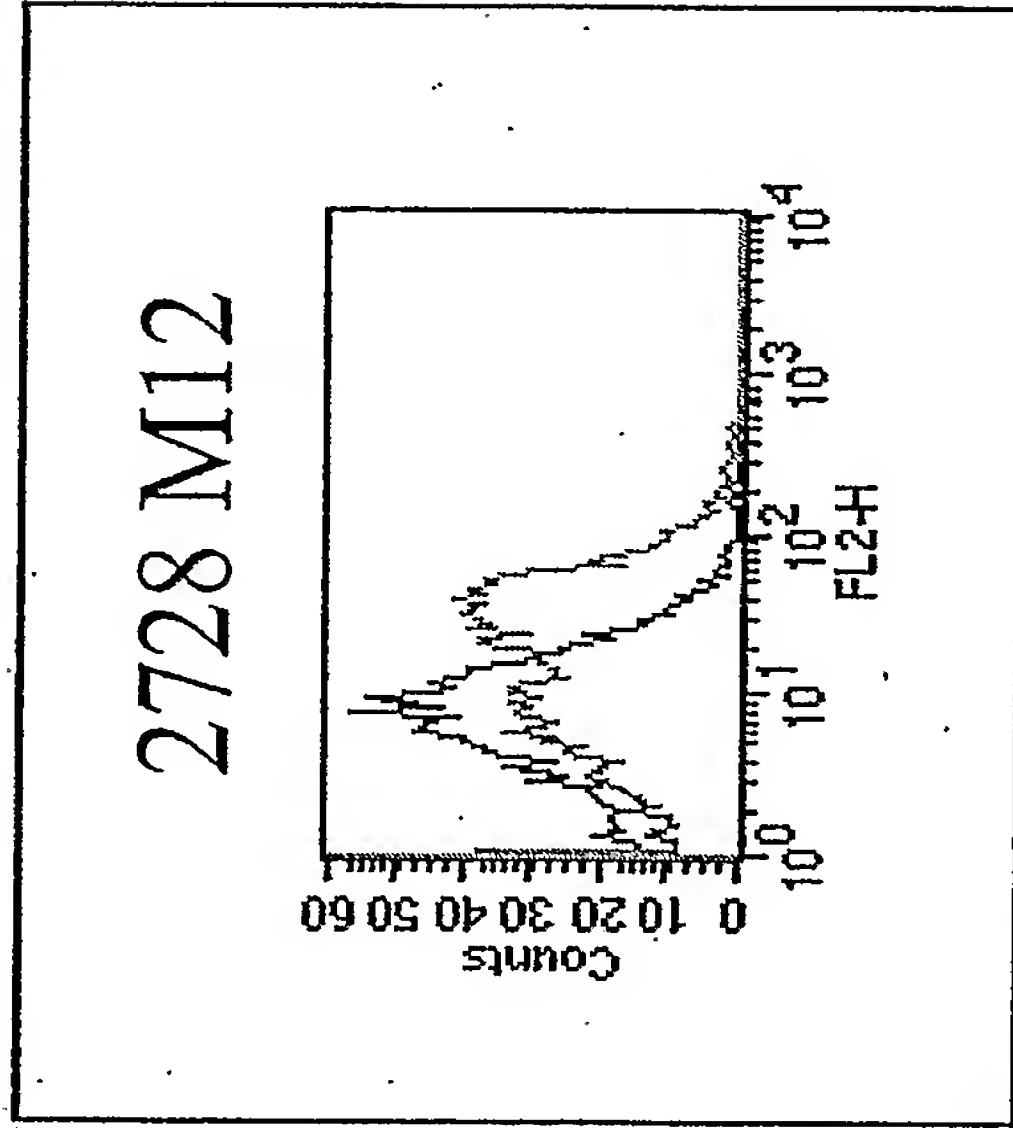
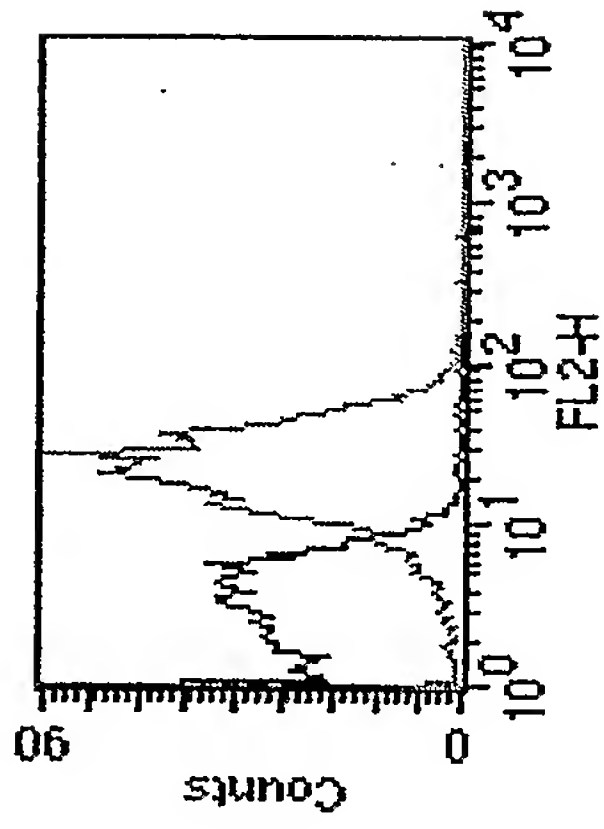
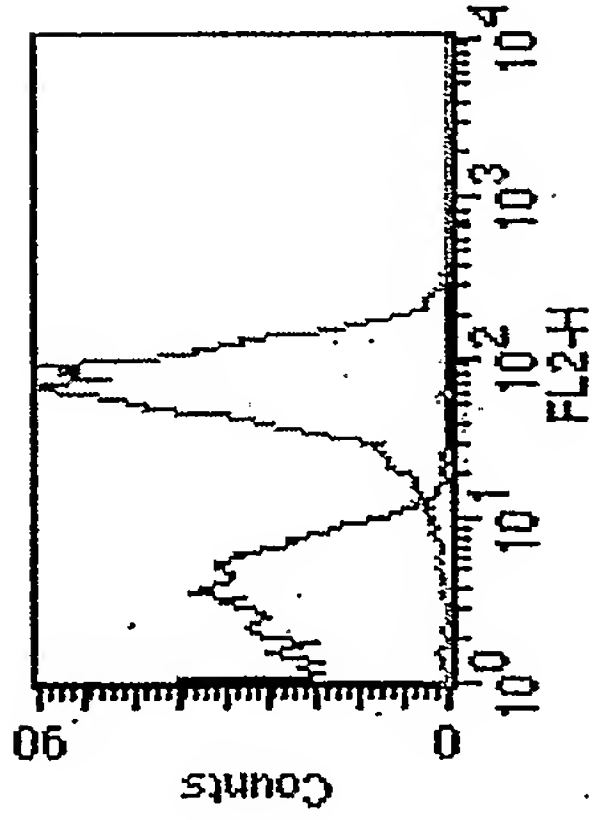


Figure 83

spyM3_0106 type 3 pilus present in M3

2721 M3



3135 M3

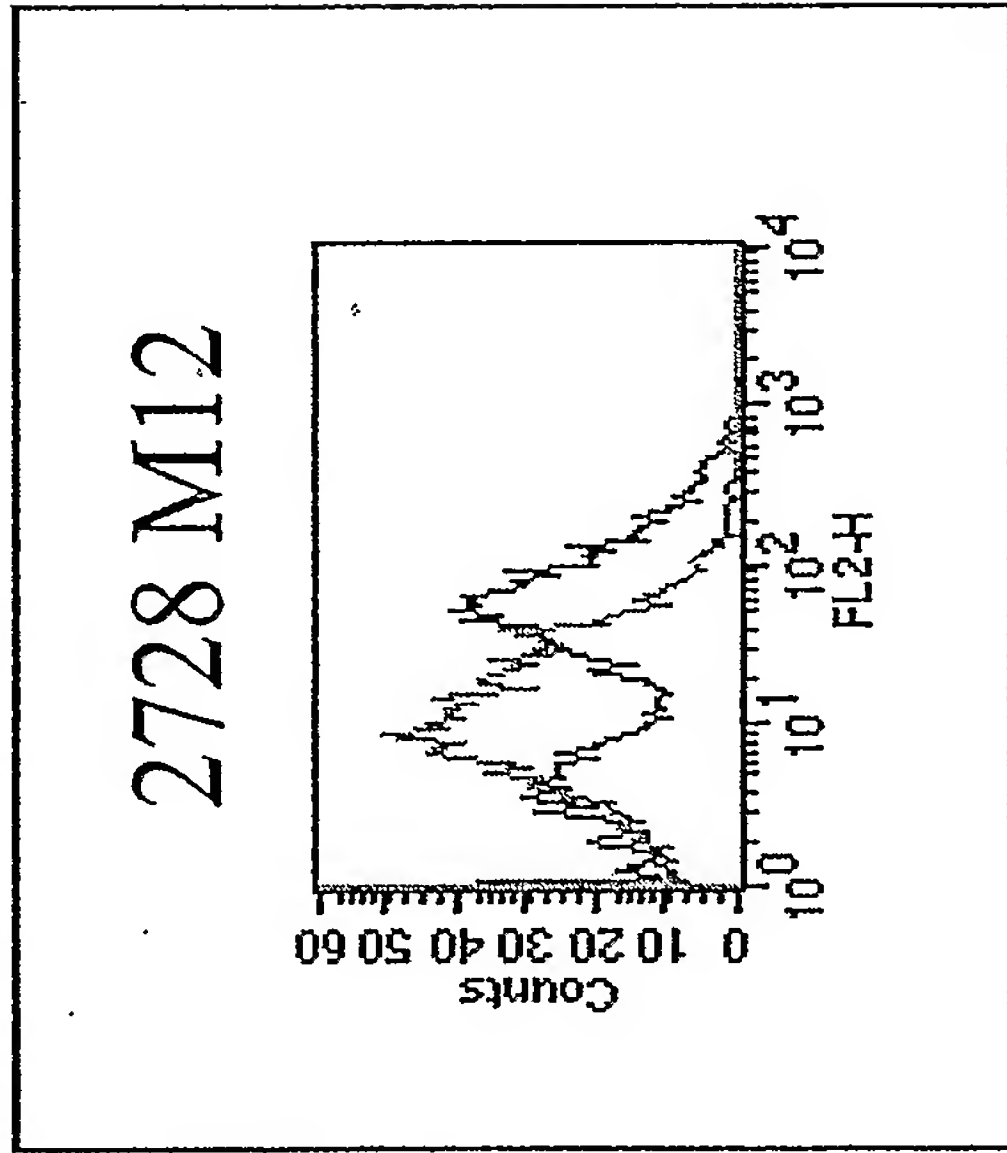
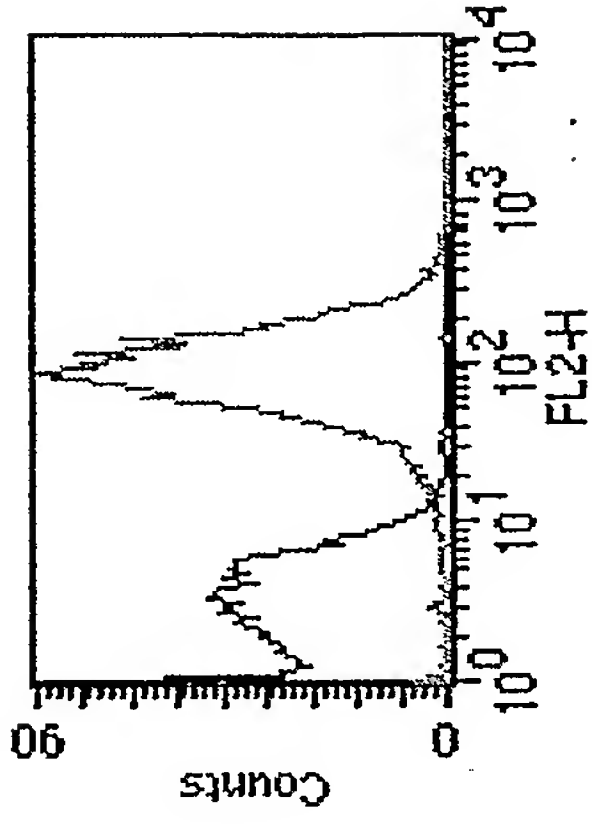
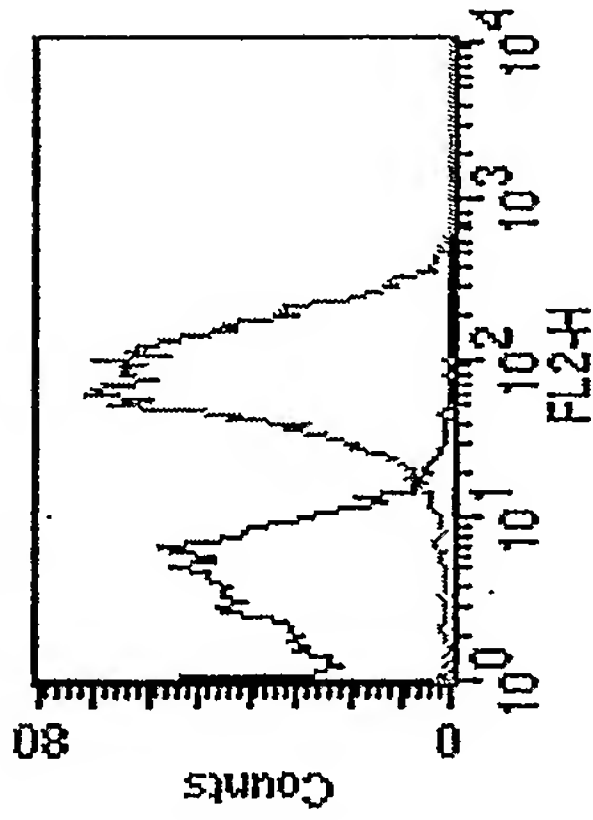


Figure 84

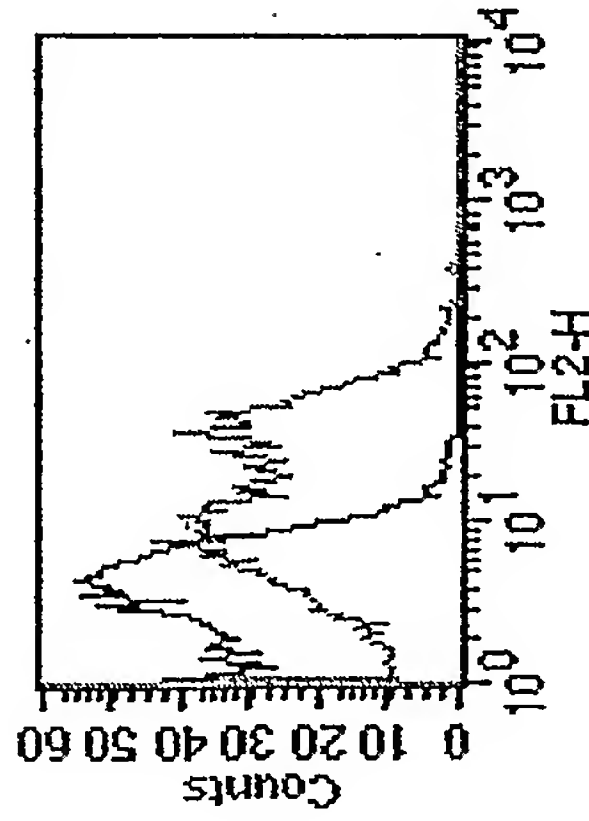
PCT/US05/27239 235/487

19224134 type 4 pilus present in M12

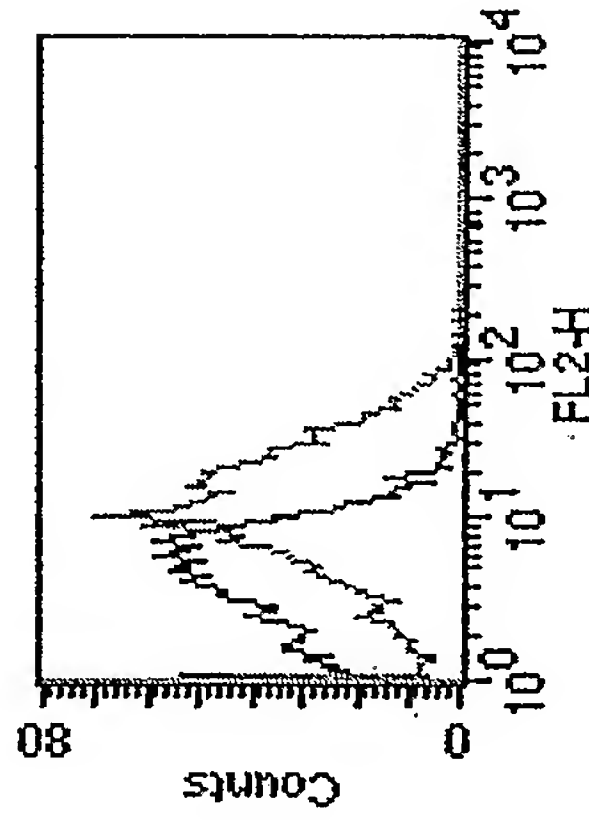
2728 M12



2724 M6



3650 M6



2894 M6

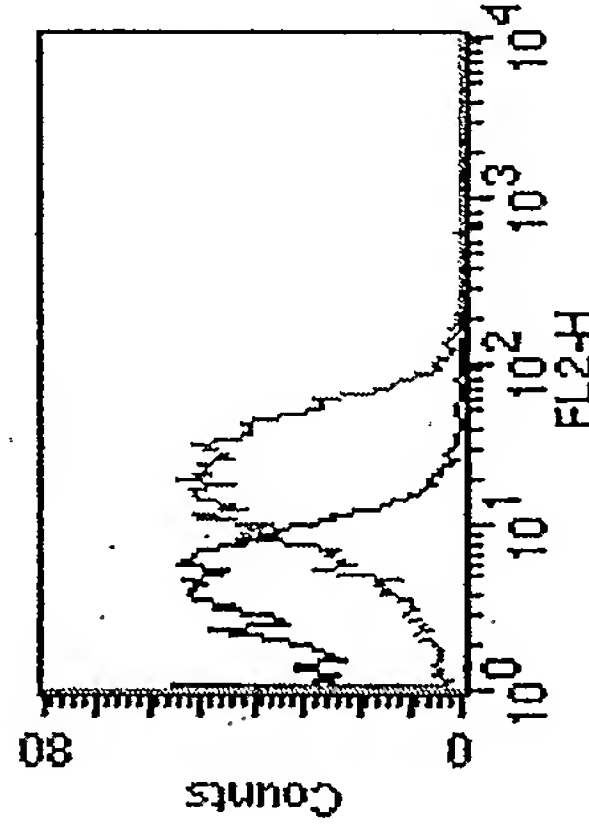


Figure 85

19224135 type 4 pilus present in M12

2728 M12

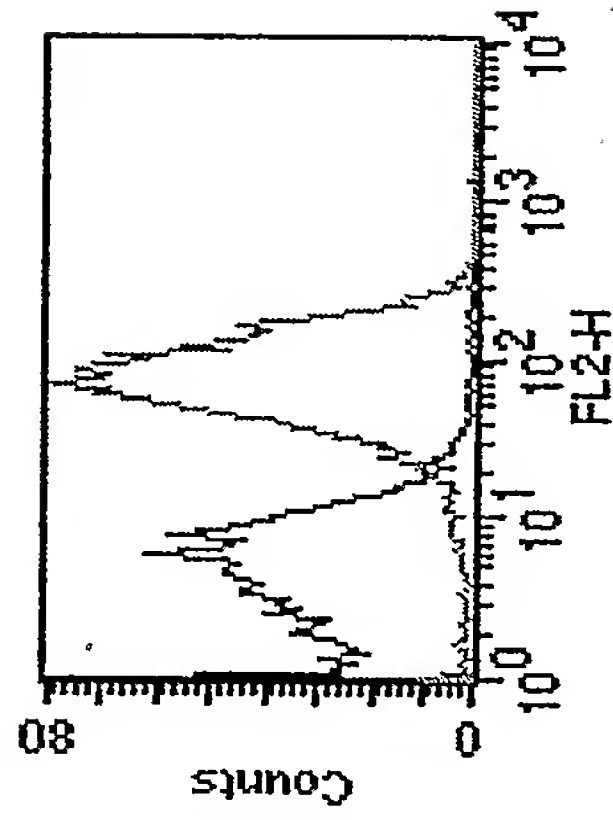


Figure 86

19224137 type 4 pilus present in M12

2728 M12

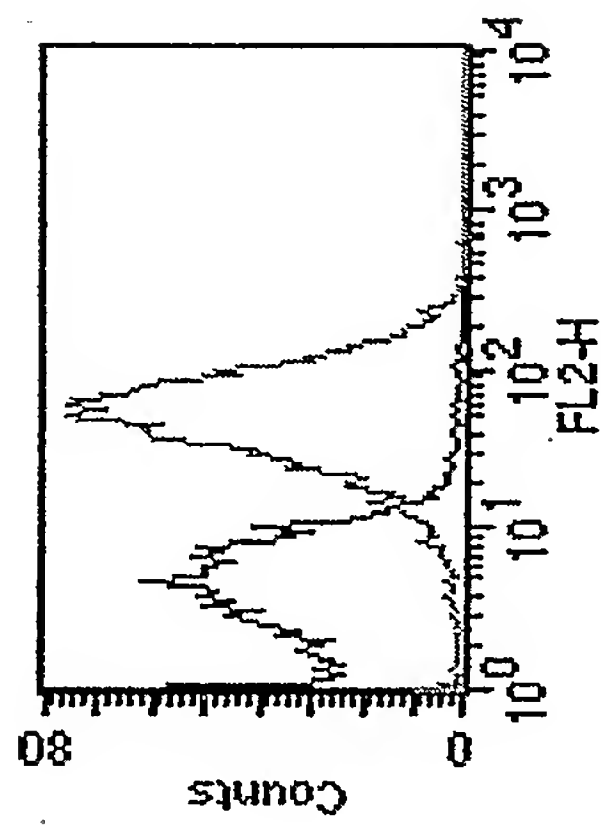


Figure 87

19224141 type 4 pilus present in M12

2728 M12

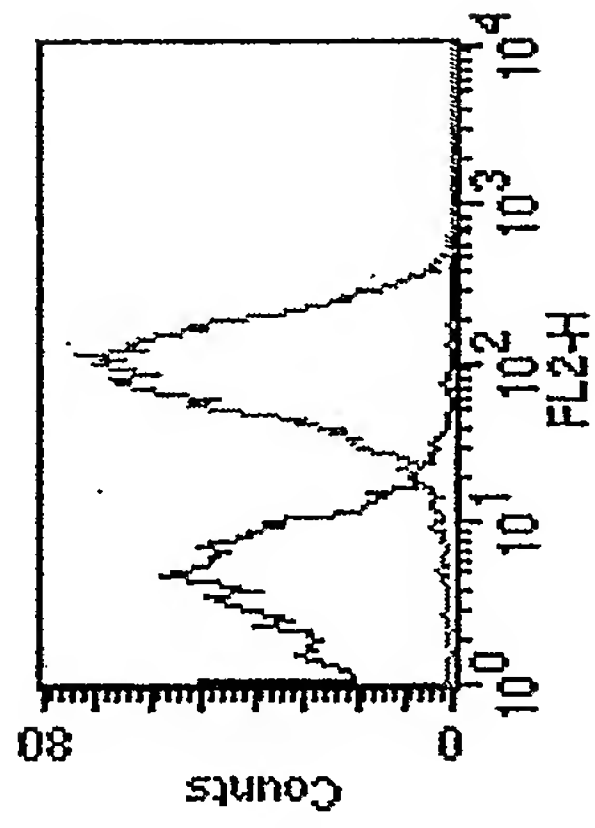
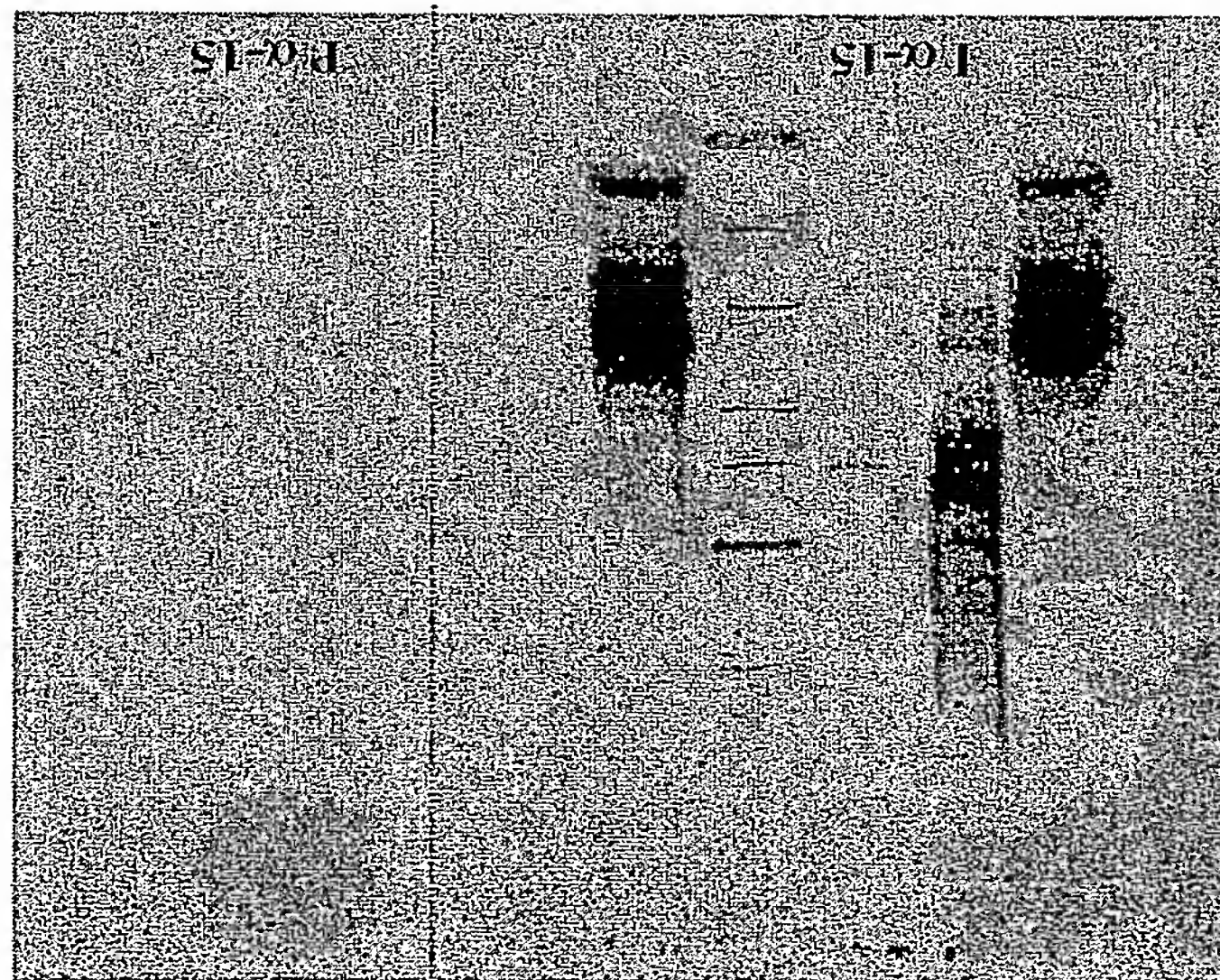


Figure 88

Figure 89



15
MI surf prot
MI tot

15
16
18

MI surf prot
MI tot

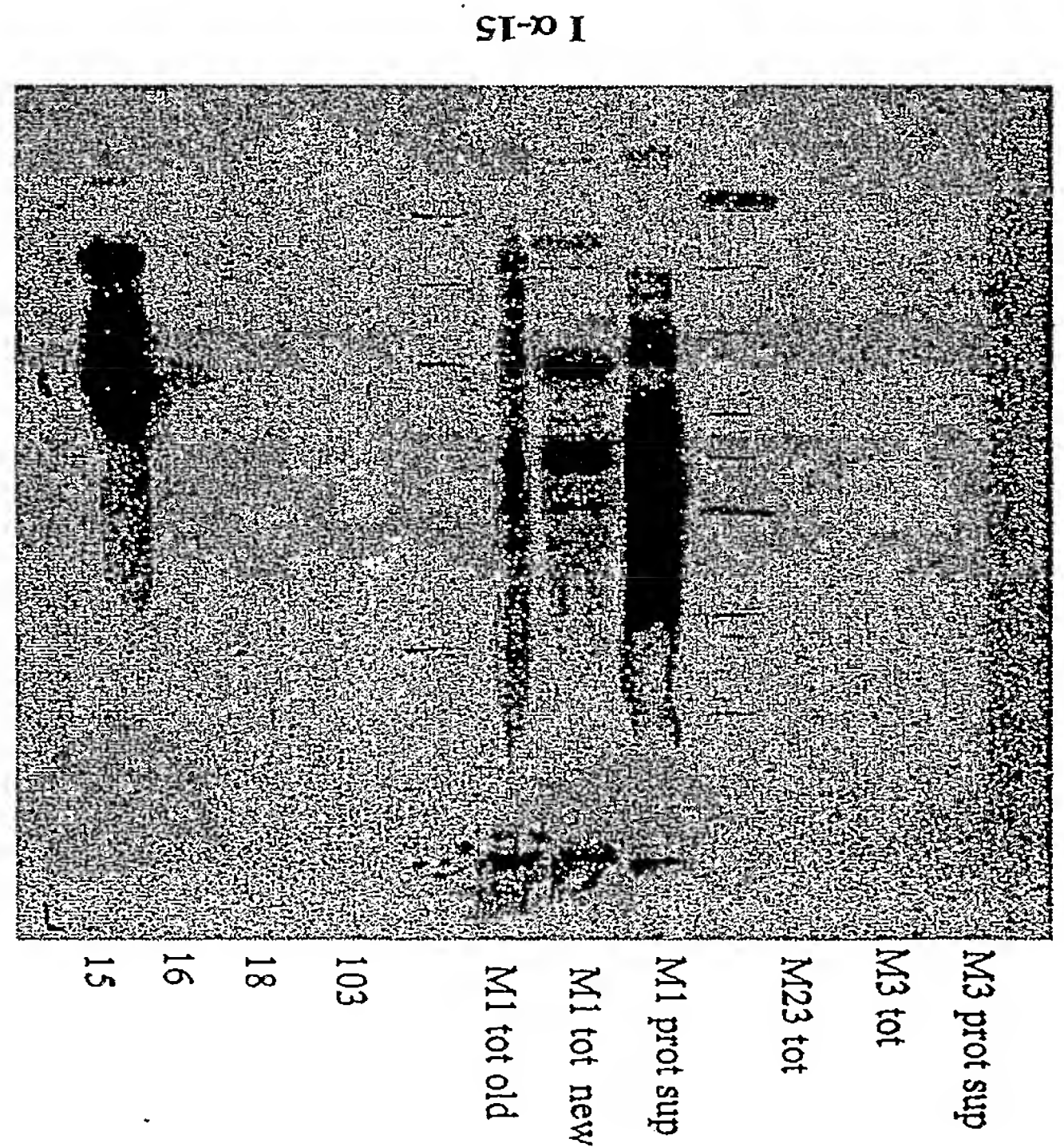
LEGEND:
 MI tot: total extract (MI)
 MI surf prot: fraction enriched in surface proteins
 #: Purified recombinant proteins, 30 ng
 I α-#: immune sera against #
 P α-#: pre-immune sera

240/487

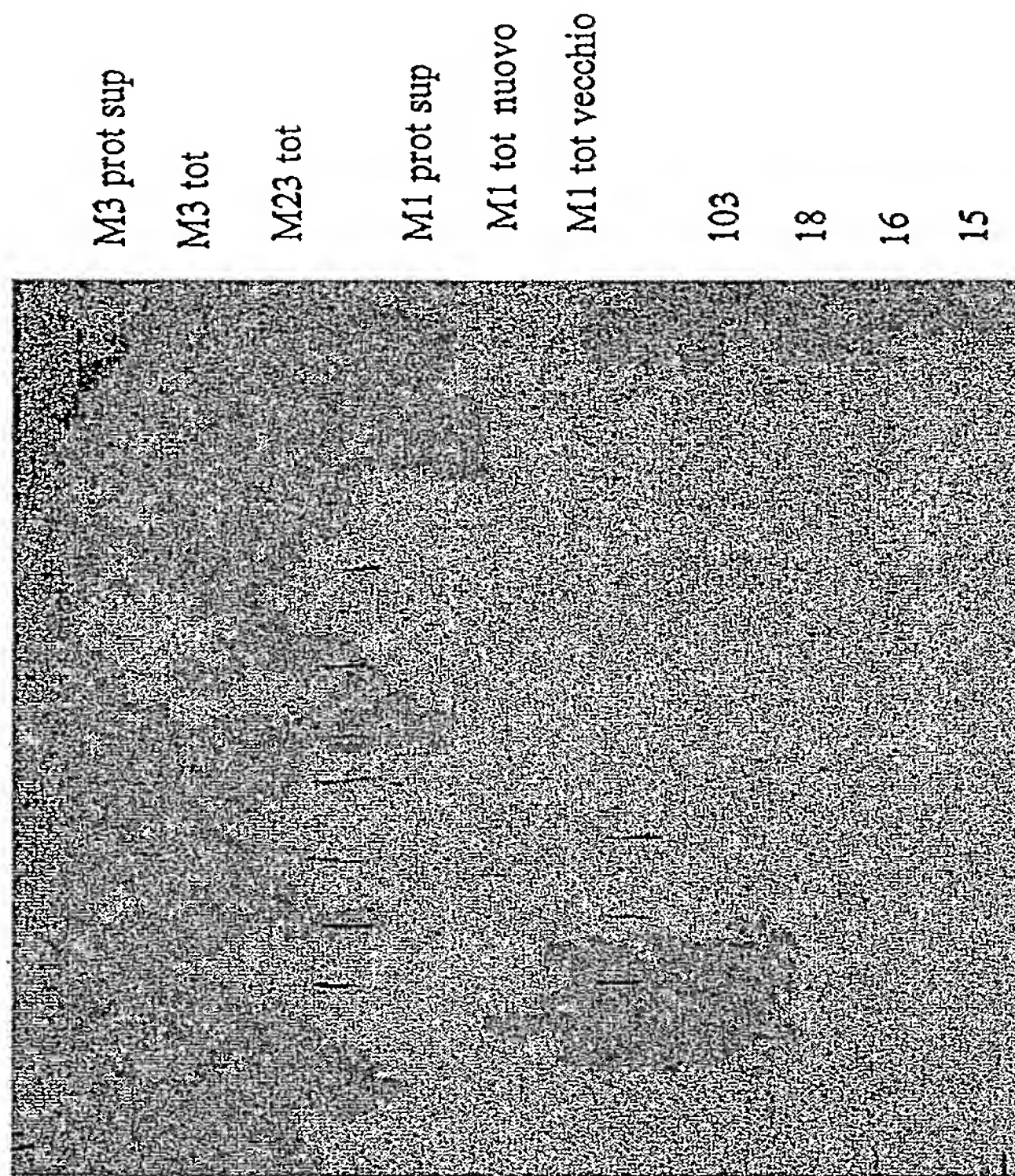
Figure 90

LEGENDA:

MI tot: total extract (MI)
 MI prot. Sup.: fraction enriched in surface proteins
 #: Purified recombinant proteins, 30 ng
 I α-#: immune sera against #
 P α-#: pre-immune sera



PCT/US05/27239 241/487

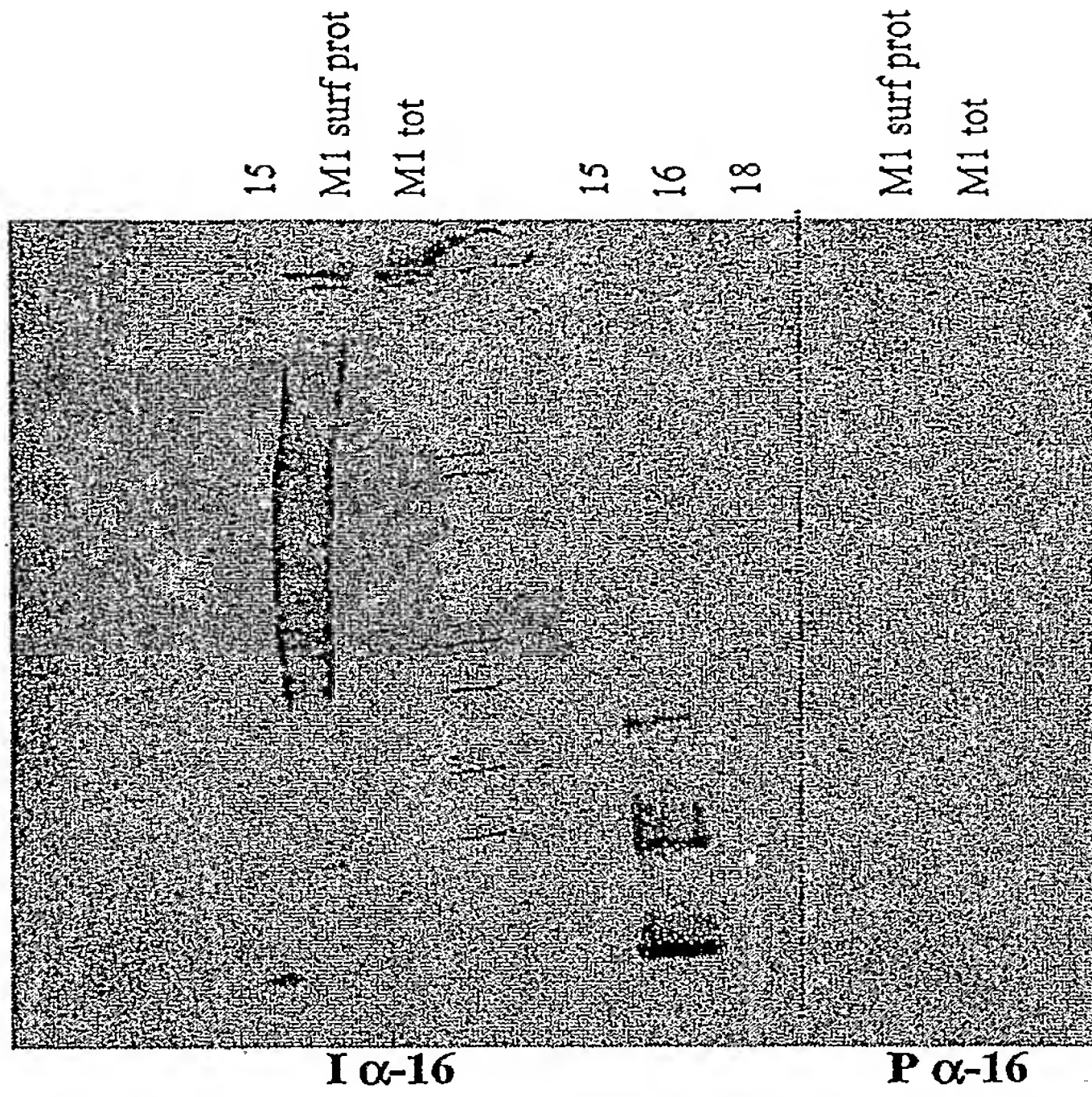


LEGENDA:
M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

P α-15

Figure 91

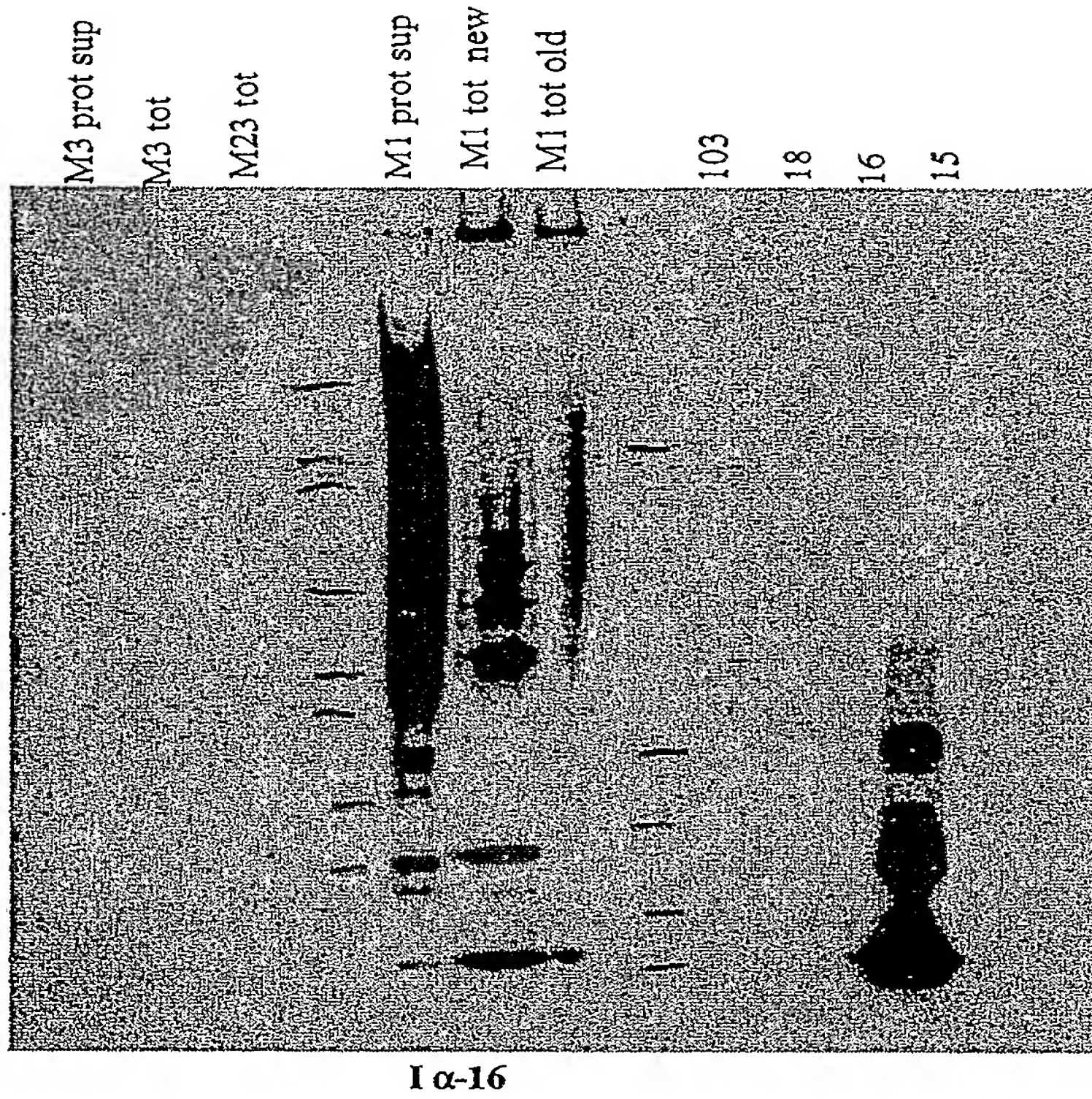
PCT/US05/27239 242/487



LEGEND:

M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

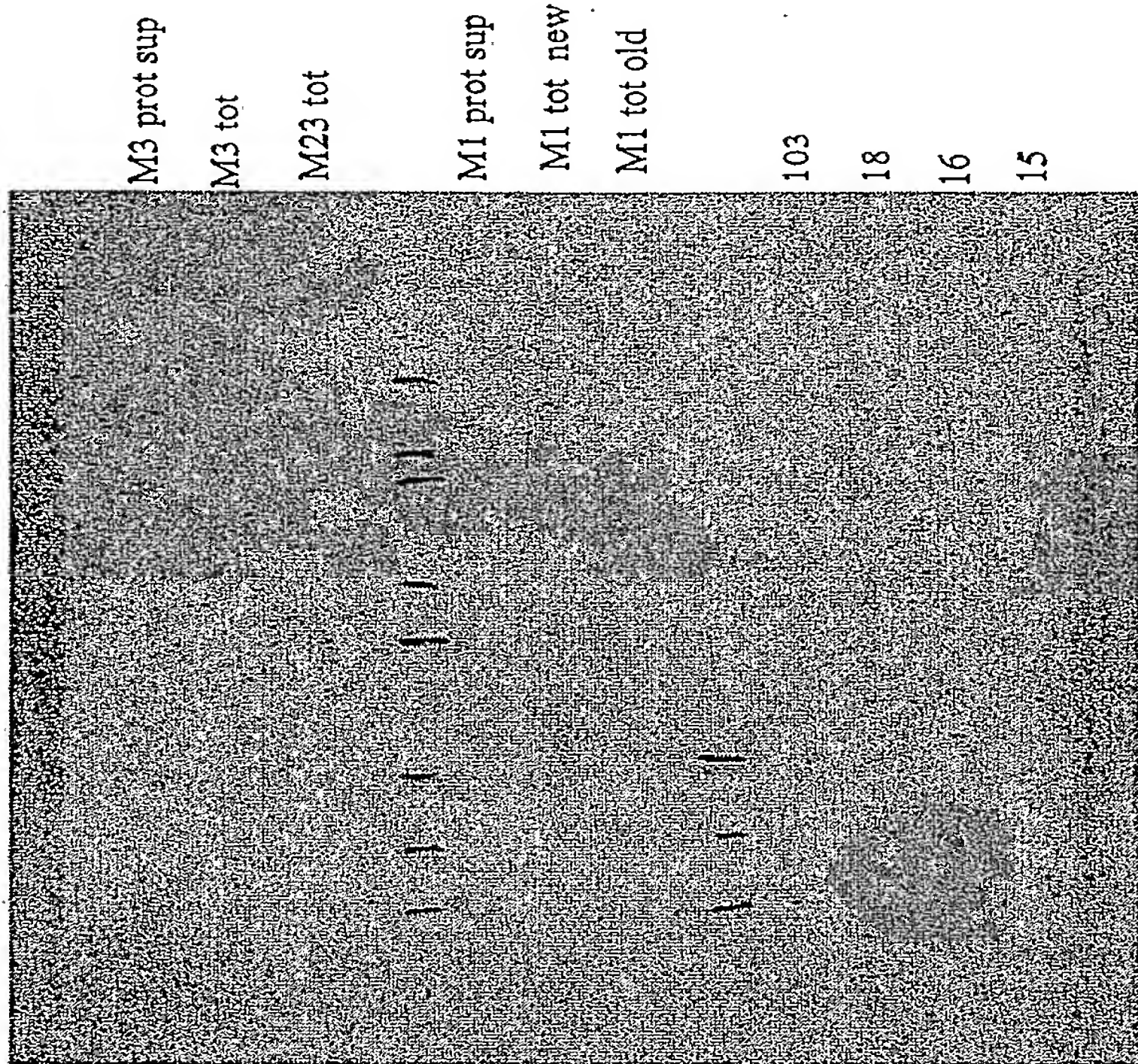
Figure 92



LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 93

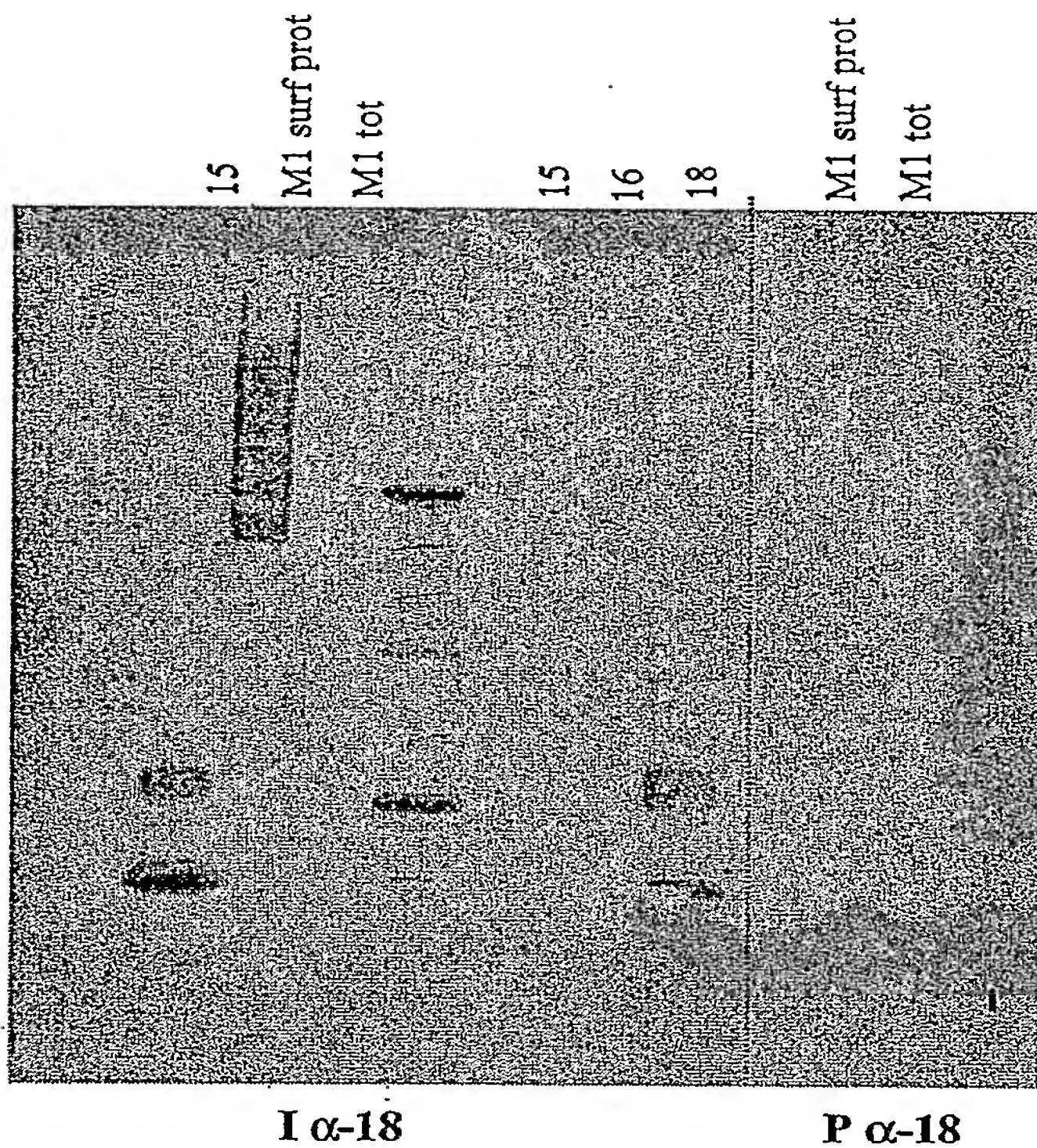


P α-16

LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 94

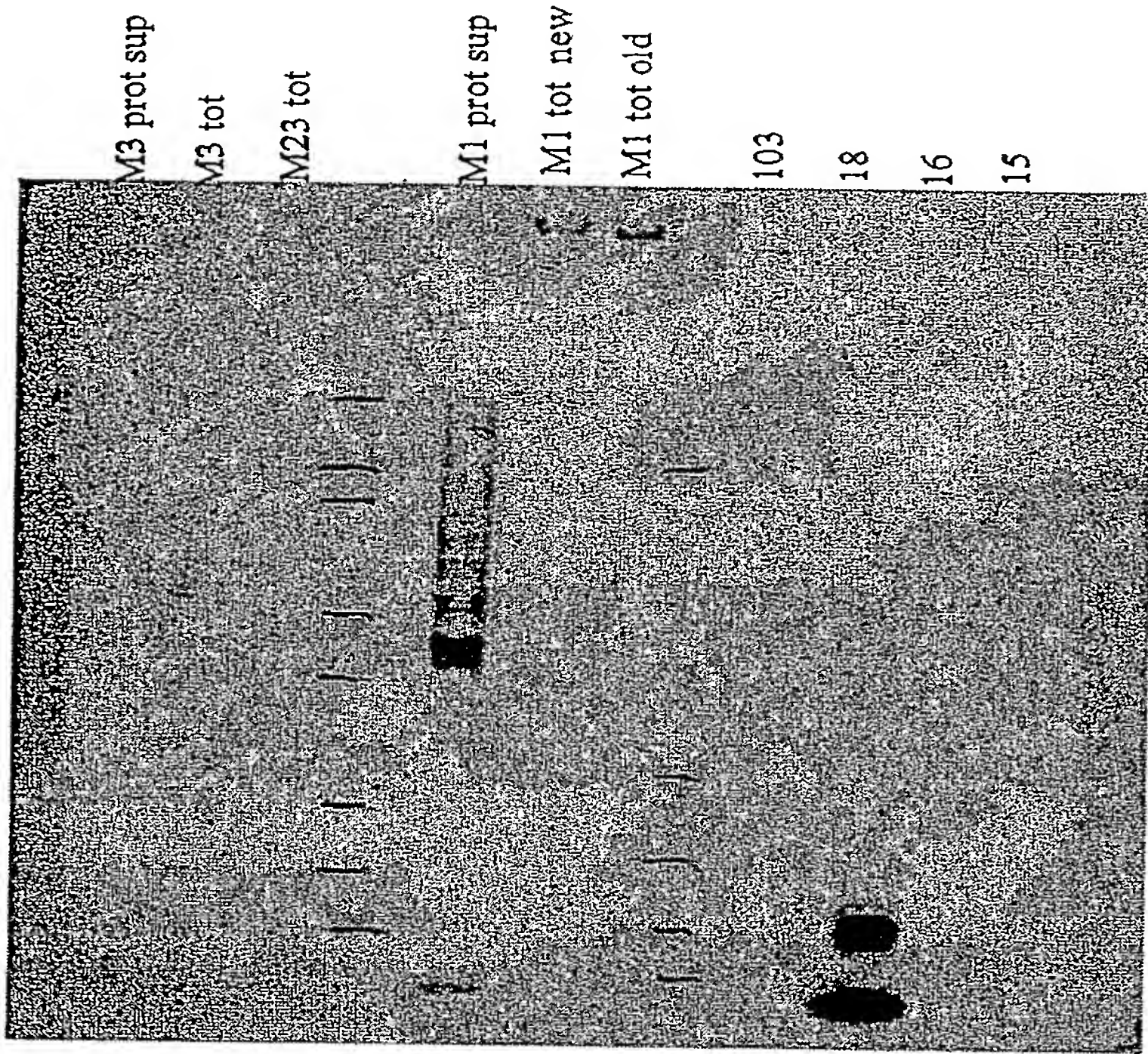


LEGEND:

M1 tot: total extract (M1)
M1 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

Figure 95

PCT/US05/27239 246/487



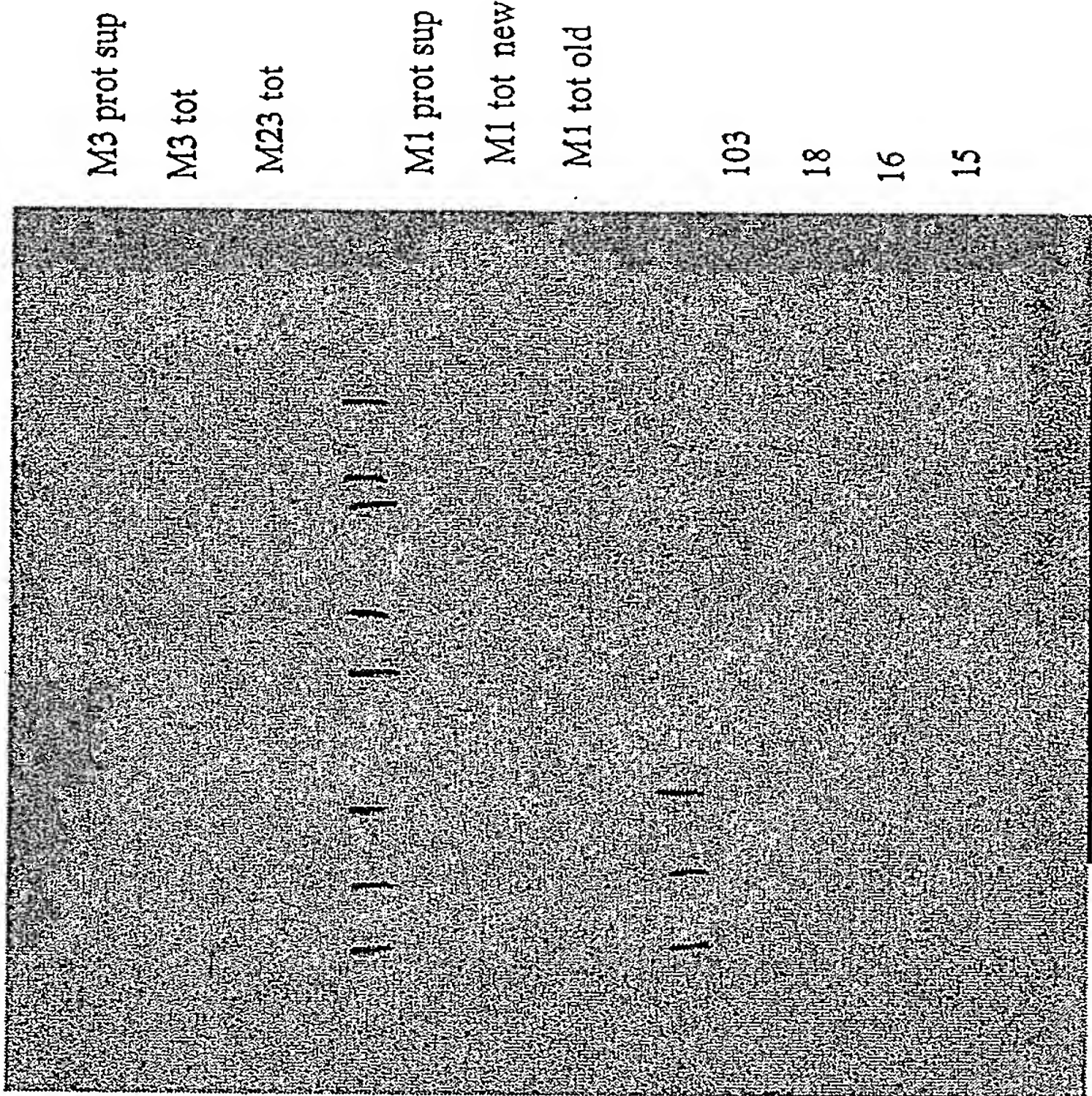
I α -18

LEGENDA:

M1 tot: total extract (M1)
M1 prot. Sup.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α -#: immune sera against #
P α -#: pre-immune sera

Figure 96

PCT/US05/27239 247/487

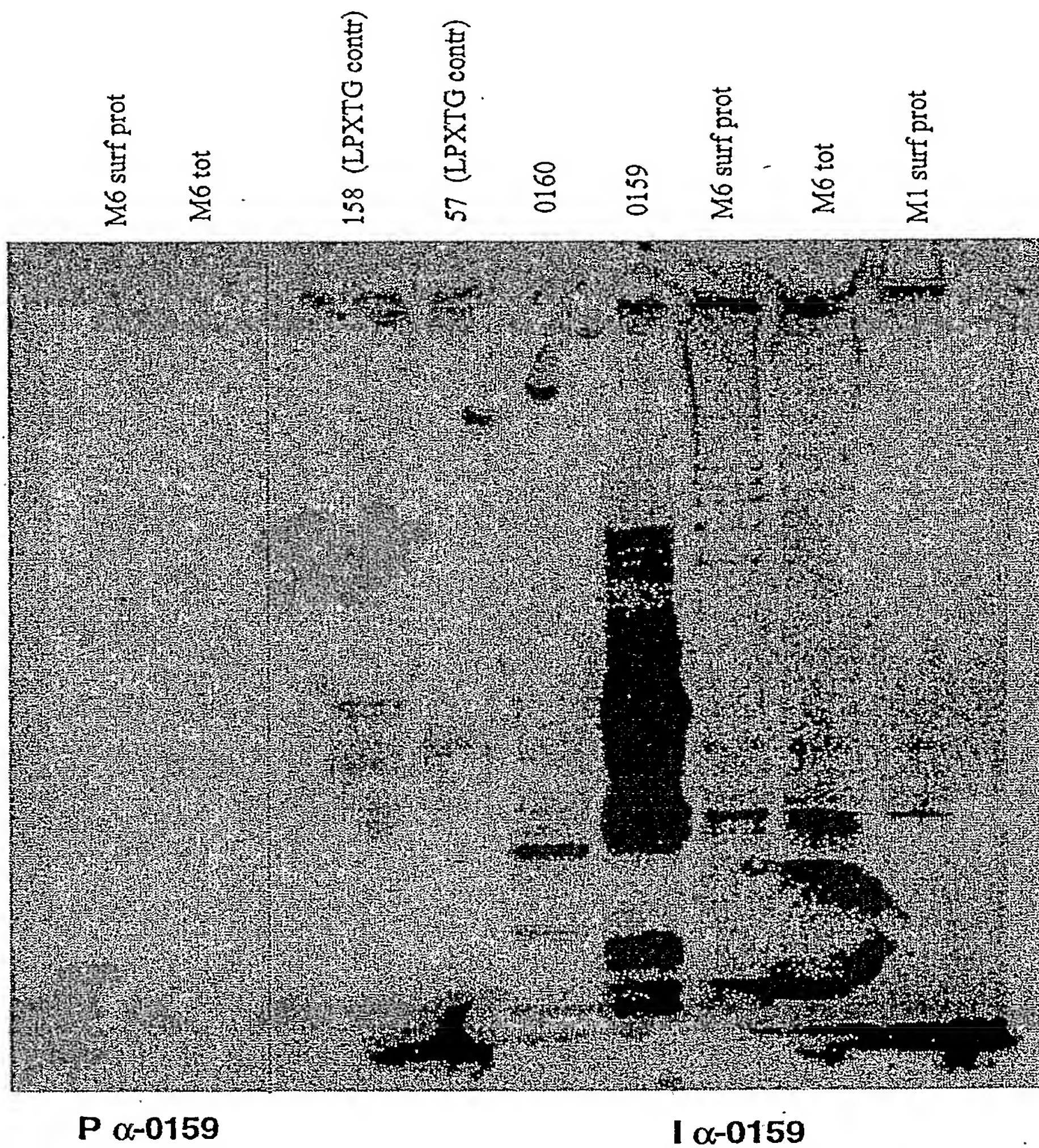


P α-18

Figure 97

PCT/US05/27239 248/487

Figure 98



LEGEND:

M6 tot: total extract (M6)

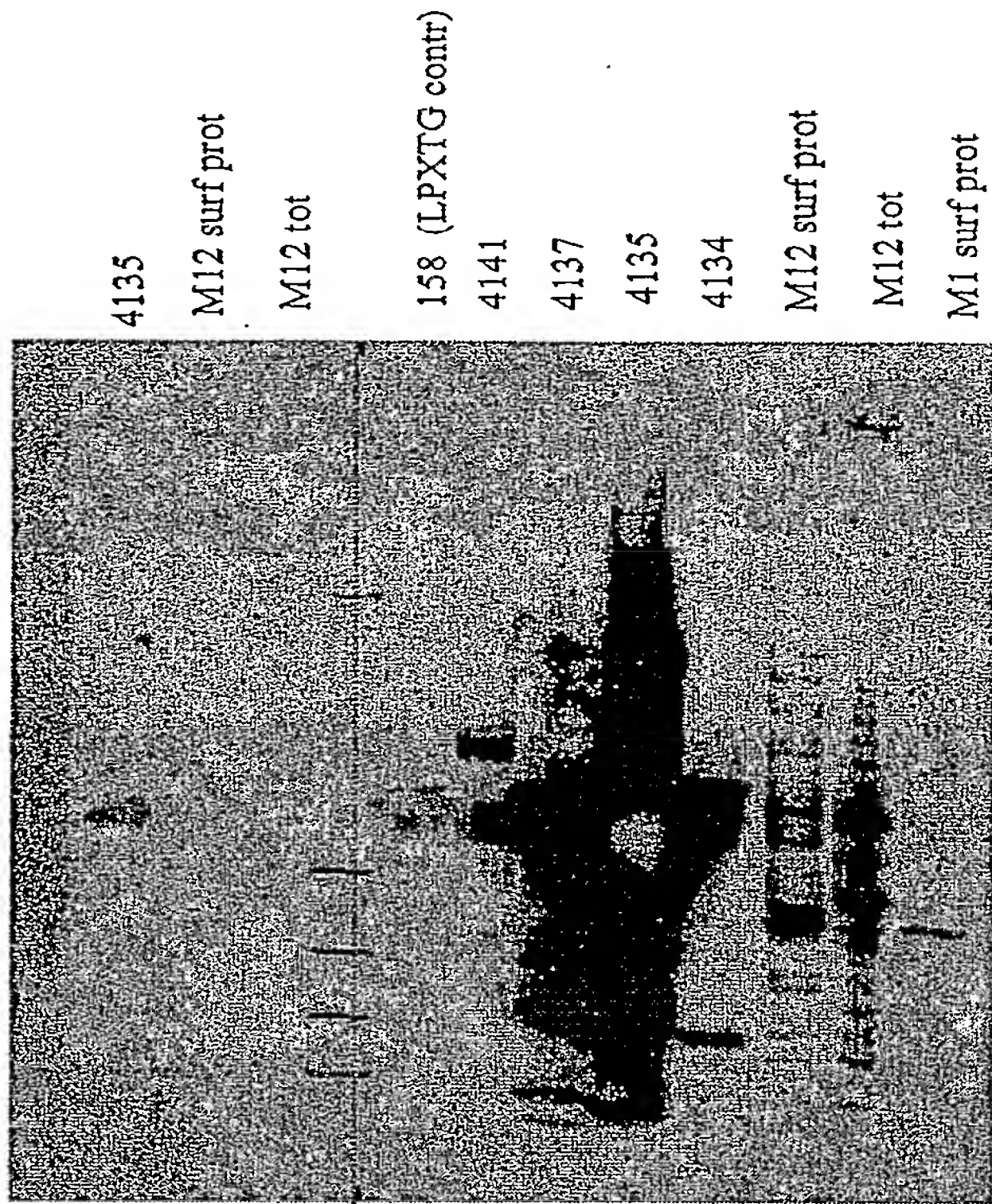
M6 surf prot.: fraction enriched in surface proteins

#: Purified recombinant proteins, 30 ng

I α-#: immune sera against #

P α-#: pre-immune sera

PCT/US05/27239 249/487



LEGEND:

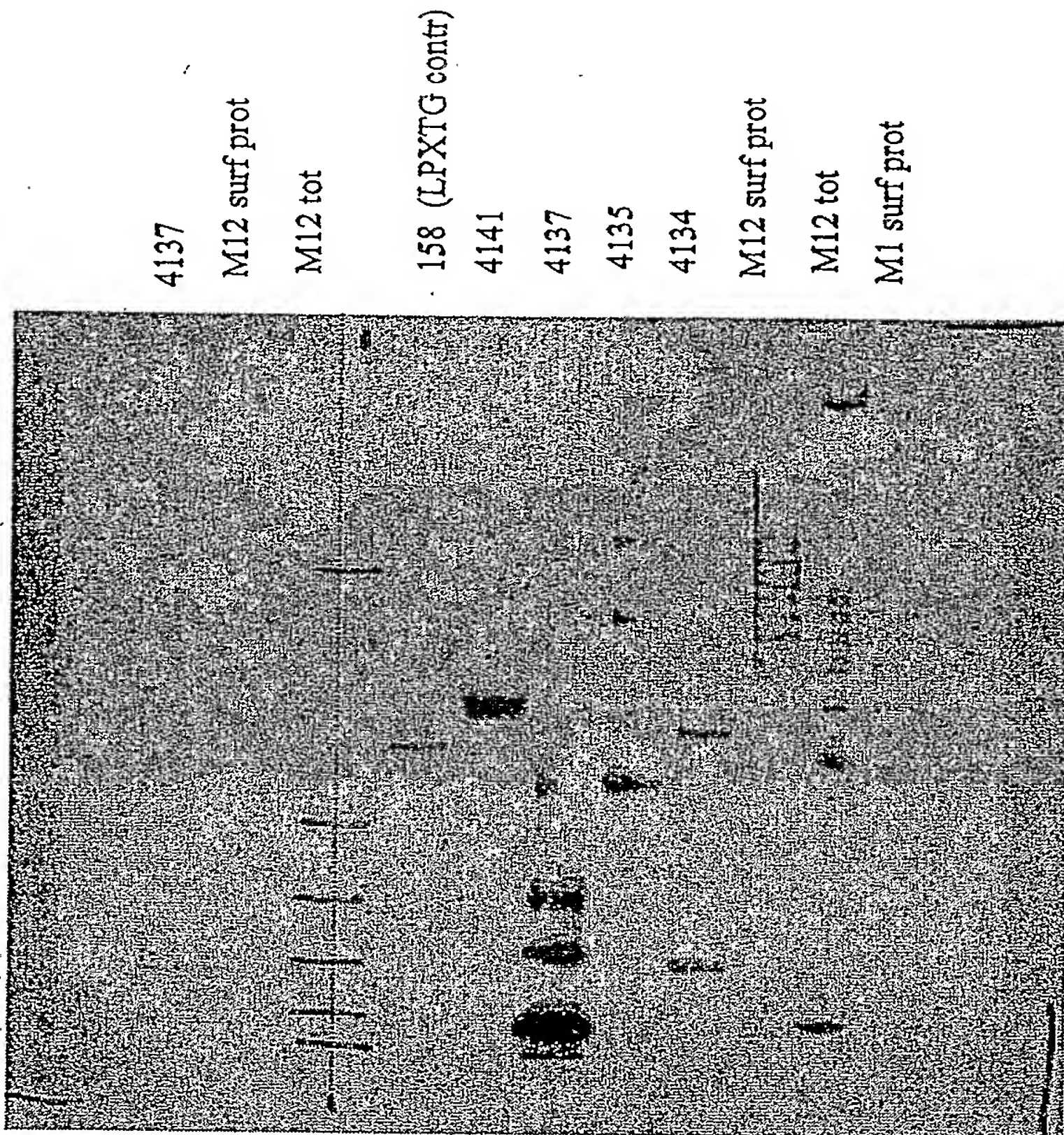
M12 tot: total extract (M12)
M12 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

P α-4135

I α-4135

Figure 99

PCT/US05/27239 250/487



LEGEND:
M12 tot: total extract (M12)
M12 surf prot.: fraction enriched in surface proteins
#: Purified recombinant proteins, 30 ng
I α-#: immune sera against #
P α-#: pre-immune sera

P α-4137

I α-4137

Figure 100

PCT/US05/27239 251/487

FIGURE 101

```

1  GACAAGCTTCCCTTATACGACCGCTTCTATATACGGACTCTTCCAAAGTTCCTGCTCTTTTA  60
   -----+-----+-----+-----+-----+-----+-----+-----+
   CTGTTCGAAGGAATATGCTGGCGAAGATATAGCCCTGAGAAGGTTCAAGAACGAGAAAAT

61  CCAGGGACTAGCCGTTCAAGGTGCAACGATGTGCGGTGTTGTTAAATGGAACCAAGTCGT
   -----+-----+-----+-----+-----+-----+-----+-----+
   GGTCCCTGATCGGCAAGTCCACGTTGCTAACAGCCACCAAAACAATTTACCTTGGTCAGCA

121  TCAGTTGTGACAGAAATTTACCCTTCTATCTTTGGGATTCGCCGTTATGTTTGGAGCTAGTGCC
   -----+-----+-----+-----+-----+-----+-----+-----+
   AGTCAACACTGTCTTAAATGGAAGATAGAACCCCTAAGGGCAATACAAACCTCGATCACGG

181  TTAAAGATTTTCAAATTTGTGAAAGCCGGGAGAACTCTTTGAGCTTTGGGCAATGTTTGTG
   -----+-----+-----+-----+-----+-----+-----+-----+
   AATTTCATAAAGTTTAAACACTTTCGGCCCTCTTTGAGAACTCGAAACCCGTTAACAAAAAC

241  CTC TTGGTCGGGATGGGAGTAGCTTTTGGCGGTCAGCATGGTGGCTATTCGCTTCTTGACC
   -----+-----+-----+-----+-----+-----+-----+-----+
   GAGAACCGCGCTACCCCTCATCGAAAACGCCAGTCGTACCACCCGATAGCGGAAGAACTGG

301  AGCTATGTGAAAAAACACGACTTCACCCCTTTTGGTAAATACCGTATCGTCTTGGTAGT
   -----+-----+-----+-----+-----+-----+-----+-----+
   TCGATACACTTTTGTGCTGAAGTGGGAAAAAACCCATTTATGGCATAGCACGAAACCATCA

361  GTTTGGCTACTTTACAGTTTTGTCCGTTTATTTGTATAAGAAAAACCTTGAAGGGTAAAC
   -----+-----+-----+-----+-----+-----+-----+-----+
   CAAAACGATGAAATGTCAAACACAGGCAATAAACAATTTCTTTTGGAACTTCCCCCATTG

421  TCTTCAAGGTTTATACCTTTAGAAAAATCTCTTCAAACCCGTCAGCTTTATCTGCAACC
   -----+-----+-----+-----+-----+-----+-----+-----+
   AGAAGTCCAAAATATGAGAAATCTTTTAGAGAAGTTTGGCCGAGTTCGAAATAGACCGTTGG

```

PCT/US05/27239 252/487

Figure 101A

481 TC AAAACAGTGT TTTGAGCAGCCCTCGGCTAGCTTCCCTAGTTTGGCTCTTTGATTTTCATT + 540
AGTTTGT CACAAAAC TCGTCGGACGCCGATCGAAGGATCA AACGAGAACTAAAAGTAA

541 GAGCTTTAAAATCCAGTCAGGGTAATCCCAATAGGGGACACCCCTCTTTCTTCTCGCTT + 600
CTCGAAAATTTAGGTCAGTCCCATTAGGGTTATCCGCCCTGTGGAGAAAGAAAGAGCGGAA

601 AATTCATATAGAGTTCAGGGCTATTTGGCTTATCTGACTAGCATCTTGTGTTTTTGG + 660
TTAAGAAGTATCTCAAGTCCCGATAAACCGAATAGACTGATCGTAGAACACAAAAAACC

661 CAAGACTTTTCGTTTGGTAAGAGTTGAAAAGTCCCTCGTAGCGGATTTTCAAAATGACAA + 720
GTTCTGAAAAGCAAACCATCTCAACTTTTCAGGAGCATCGCCATAAAGTTTTTACTGTT

721 TTTTCCAGCTTTTCTTGTGATGATGATGAGAGCGACTTTTTCGTGATAGAAGAGTCA + 780
AAAAAGT CGAAAAGAACAACTACATCTA ACTCTCGCTGAAAAGACTATCTTCTCAGT

781 GCTCTTTTGTGATATCTTCCCTCGGCACGGAGAACTTCCCGTAGGTTTTCTCCTTGCCGA + 840
CGAGAAAAC TATAGAGGAGCCCTCTTTAGAGGGCATCCAAAAGAGGAACGGCT

841 TTGATTTACGGATCGGATTTGGATTTGACTGGAGAGTTGTGAATGCCACGAGCCTTTCGAT + 900
AACTAAATGCCCTACGCTAACCTAAACTGACCTCTCAACACTTACGGTCTCGGAAAAGCTA

901 ACAGATCATAGCCTAGTCTACCAAAACGGTCTATTAGGGTTACCTCAGGAACITCAAGTA + 960
TGTC TAGTATCGGATCAGATGGTTTTGGCCAGATAATCCCAATGGAGTCCCTTGAAGTTCAT

PCT/US05/27239

253/487

Figure 101B

961 AATCAGCACAGTAAACCCATTTGATGAAGACGTTCTACTGTCTTTTCCCTACTC 1020
 -----+-----+-----+-----+-----+-----+-----+
 TTAGTCGTGGTCAATTTGGGGTAAACTACTCTCTGCAAGATGACAGAAAAAAGGATGAG

 1021 CATGAAATTTGGAATATCCATTTGTTGAGAAAAATCCCTCAGCCCTGTTCCAGGTAGAATCA 1080
 -----+-----+-----+-----+-----+-----+-----+
 GTACTTTAAACCTTTATAGGTAACAACAACCTCTTTTAGGAGTCGGACAAAGTCCCATCTTAGT

 1081 CTGTCAAACCATGTTGTTGATAATCACTCGCCATTTTAGCTAAGAATTTGTTGTAAG 1140
 -----+-----+-----+-----+-----+-----+-----+
 GACAGTTTGGTACACCAAAACTAATTAGTGAGCGGTAATAATCGATCTTTAAACAACATTC

 1141 AAACGCCCTGGGAAAGCAGTTAGATGGAGTCTTTCCAGATACTTTTGAATGAGCGGAG 1200
 -----+-----+-----+-----+-----+-----+-----+
 TTTGGGACGCCCTTCGTCAATCTACCTCAAGAAAGGTCTATAGAAAAAACTTACTCCGCTC

 1201 CAATTTGACCCGTGACTTGATACCGAGTTTATTTCTGTCAACATCCAAATAGGCCTTCGT 1260
 -----+-----+-----+-----+-----+-----+-----+
 GTTAAACCTGGGACTGAACATATGGCTCAATAAAGACAGTGTAGGTTTATCCGAAGCA

 1261 CAATGCTCATGGGTTCAATCAATCTGTATAGCGCTTAAAAAATAGCTCGAATCCGGAGTC 1320
 -----+-----+-----+-----+-----+-----+-----+
 GTTACGAGTACCCAAAGTTAGTTTAGACATATCGCGAATTTTATCGAGCTTAGGCCCTCAG

 1321 CCACAGACTTGTATTTCTCATAATCCCTGAGATAAAGACAGCCCTGGGGACAACGTTTCAT 1380
 -----+-----+-----+-----+-----+-----+-----+
 GGTGCTGAACATAAAGAGTATTAAGGGACTCTATTTCTGTCTGGACCCCTGTTGCAAGTA

 1381 AAGCTTCCCTTGGAACTCATGGCAGAAATGGACACCAAAAGCTCTTGCCTCATAACTACAGG 1440
 -----+-----+-----+-----+-----+-----+-----+
 TTCGAAGGAACCTTGAGTACCGCTTACCCTGTGGTTTTTCGAGAACGGAGTATGATGTC

PCT/US05/27239 254/487

Figure 101C

1441 TAGAAACGACTCCCCCGTCCACCTGTTTGGCCGAGGGTCGTTCCAAATAATGACAGGTTTTC
-----+-----+-----+-----+-----+-----+-----+
1500 ATCTTTGCTGAGGGCCAGGTGGACAAACGGCTCCAGCGAAGGTATTACTGTCCAAAAG

1501 CTCTGAGTTTAGGATTATCCCTGATTTCCACTGCAGCAAAAAGGCATCCATGTCAATAT
-----+-----+-----+-----+-----+-----+-----+
1560 GAGACTCAAATCCTAATAGGGACTAAAGGTGACGTCGTTTTCCTCCGTAAGGTACAGTTATA

1561 GGATGATTTTCTTGACAAATCATTTAACAAGGAAAATCAACATGCCCTAGCACCTTTT
-----+-----+-----+-----+-----+-----+-----+
1620 CCTACTAAAAGAACTGTTTAGTAAATTTGTTTCCCTTTTAGTTGTACGGATCCGTGGAAAA

1621 TATACTCTTCGAAAATCTCTTCAACCACGTCAGTTCCTCATCTGCAACCTCAAAAACAGTA
-----+-----+-----+-----+-----+-----+-----+
1680 ATATGAGAAGCTTTTAGAGAAGTTTGGTGCAGTCCGAAGGTAGACGTTTGGAGTTTGTTCAT

1681 TTTTGAGCTGACTTCGTCAGTTCATTTACAACCTCAAAGCAGTGTCTTTGAGCAGCCGTC
-----+-----+-----+-----+-----+-----+-----+
1740 AAAACTCGACTGAAGCAGTCAAGATAAATGTTGGAGTTTCGTCACGAAACTCGTCGGGACG

1741 GGCTAGTTTCCTAGTTTGGCTTTTCGATTTCCATTTGAGTGTAACTGCTTATTTCTTTTAT
-----+-----+-----+-----+-----+-----+-----+
1800 CCGATCAAAGGATCAAACGAAAAGCTAAAGGTAACCTCACATTTGACGAATAAAGAAAATA

1801 TATACCTTTTTCGAAAAAAGAAAAGGACTTTATTTTTCAAAAATAATAACA
-----+-----+-----+-----+-----+-----+-----+
1860 ATATGGGAAAAAAGACTTTTTCCTTTTCTTTTTCCTGAAAATAAAGTTTATATATATGT

1861 GTTTGAAAATAAATAGACTGTTTGTAGAAAAGAAAGTGTAAAATAAGGAATTTTTCACCT
-----+-----+-----+-----+-----+-----+-----+
1920 CAAACTTATTTTATATCTGACAAAATCTTTTCTTTTCTTTTTCACATTTTATCCTTAAAGAAGTGA

Figure 101D

1921 TGTTGAAAATCGGTTACTTTATGGTATACTTGTCTCAATGAATGTAACAGATGACTGTTACT + 1980
 ACAACTTTAGCCCAATGAAATACCATATGAACAGAGTACTTACATTTGTCTACTGACAATGA

1981 AGAAAAAGAGGACATTAATATGGTTGTTAAGACAGTTGTTGAAGCACAAAGATATTTTGG + 2040
 TCTTTTTCCTCGTGAATATATACCAACAATTCGTCAACAACCTTCGTGTTCTATATAAAAAC

c M V V K T V V E A Q D I F D - this orf is the homologue of sp0459, a formate
 acetyltransferase (pfl). It is out of the
 pilus locus

2041 ACAAAAGCTGGGAAGGCTTCAAAGCGTAGATTGGAAGAAAAGCAAGTATCACCGCT + 2100
 TGTTTCGAACCCCTTCGGAAGTTCCGCATCTAACCTTCTTTTTCGTTCCACATAGTGGGA

c K A W E G F K G V D W K E K A S V S R F -

2101 TTGTACAAGCTAACTACACACCTTATGATGGAGACGAAAGCTTCCTTGCAGGACCAACAG + 2160
 AACATGTTCCGATGATGTTGGAAATACCTACCTCTGCTTTCGAAAGGAACGTCCTGGTTGTC

c V Q A N Y T P Y D G D E S F L A G P T E -

2161 AGCGTTCACATCAAGAAAATTGTAGAAGAACTAAAGCACACTACGAAGAAACTC + 2220
 TCGCAAGTGAAGTGTAGTTCTTTTAAACATCTTCTTTGATTTTCGTGTGATGCTTCTTTTGAG

c R S L H I K K I V E E T K A H Y E E T R -

2221 GTTTCCCAATGGACACTCGTCCACATCTATCGCTGATATCCCTGCTGGATTTATCGACA + 2280
 CAAAGGTTACCCTGTGAGCAGGTTGTAGATAGCGACTATAGGACGACCTAAATAGCTGT

c F P M D T R P T S I A D I P A G F I D K -

2281 AAGAAAATGAAGTTATCTTTGGTATCCAAAATGATGAACCTCTCAAATGAACTTCATGCC + 2340
 TTCTTTTACTTCAATAGAAAACCATAGGTTTACTACTTGTAGAAAGTTTAACTTGAAGTACG

Figure 101E

C E N E V I F G I Q N D E L F K L N F M P -
 CAAAAGGTGGTATCCGTATGGCTGAAACTACTTTAAAAGAAAATGGATACGAACCAGACC 2400
 GTTTCCACCATAGGCATACCGACTTGGATGAAATTTCTTTTACCTATGCTTGGTCTGG
 C K G G I R M A E T T L K E N G Y E P D P -
 CAGCTGTTCCAGAAAATCTTCACTAAATATGTAACAACAGTTAACGACGGTATTTCCGGTG 2460
 GTCGACAAGTGCCTTTAGAAAGTATATACATTTGTTGTCATTTGCTGCCATAAAAGGCAC
 A V H E I F T K Y V T T V N D G I F R A -
 CCTACACTCAAAATATTCGTCGGCTCGTCAATGCACACACTGTAACCTGGTCTCCAGATG 2520
 GGATGTGAAGTTTATAAGCAGCGGAGCAGTACGTGTGTGACATTGACCAGAAGGTCCTAC
 Y T S N I R R A R H A H T V T G L P D A -
 CATACTACGGGACGTATCATCGGTGTTTACGCACGTCCTTCTTTACGGTGCAGACT 2580
 GTATGAGTGGCCCTGCATAGTACCCACAAAATGCGTGCAGAACGAGAAAATGCCACGCTGGA
 Y S R G R I I G V Y A R L A L Y G A D Y -
 ACTTGATGCAAGAAAAGTAAACGACTGGAAATGCAATCAAGAAAATCGATGAAGAAACAA 2640
 TGAACCTACGTTCTTTTTCATTTGCTGACCTTACGTTAGTTTCTTTTAGCTACTTCTTTGTT
 L M Q E K V N D W N A I K E I D E E T I -
 TCCGTCCTTCGTTGAAGAAGTAAACCTTCAATACCAAGCATTTGCAACAAGTTGTTCCGCCCTGG 2700
 AGGCAGAAGCAGCTTCTTTCATTTGGAAGTTATGGTTGTTAAGCTTGTTCACAACAAGCGGACC
 R L R E E V N L Q Y Q A L Q Q V V R L G -
 GTGACCTTTACGGGGTTGATGTTCCGCAAAACCAGCGATGAACGTGAAAGAAAGCAATCCAAT 2760
 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

PCT/US05/27239 257/487

Figure 101F

CACTGGAAATGCCCAACTACAAGCGTTTGGTCGCTACTTGCACCTTCTCGTTAGGTTA
 D L Y G V D V R K P A M N V K E A I Q W -
 GGGTTAACATTCCTTTCATGGCTGTCGCCGTGATTAACGGTGCCTGCTACATCTCTAG
 2761 -----+-----+-----+-----+-----+-----+-----+-----+ 2820
 CCCAATTGTAACGAAAGTACCAGACAGACGGCACACTAATGCCACGACGATGTAGAGATC
 V N I A F M A V C R V I N G A A T S L G -
 GTCGTGTACCAATCGTATTGGACATCTTTGCAGAACGGTGCCTTGCCTGGTACATTTA
 2821 -----+-----+-----+-----+-----+-----+-----+ 2880
 CAGCACATGGTTAGCATAACTGTAGAAAACGCTTTCACACTGGAACGACCCATGTAAAT
 R V P I V L D I F A E R D L A R G T F T -
 CTGAATCAGAAAATCCAAGAAATTCGTTGATGATTTTCGTTATGAAAACCTTCGTACAGTTAAAT
 2881 -----+-----+-----+-----+-----+-----+-----+ 2940
 GACTTAGTCTTTAGGTTCTTAAGCAACTACTAAAGCAATACTTTGAAGCATGTCAATTTA
 E S E I Q E F V D D F V M K L R T V K F -
 TTGCTCGTACCAAGCTTATGACCAATGTACTCAGGTGACCCAACTTTATCACAACCTT
 2941 -----+-----+-----+-----+-----+-----+-----+ 3000
 AACGAGCATGGTTTCGAATACTGGTTAACATGAGTCCACTGGGTTGGAATAAGTTGAA
 A R T K A Y D Q L Y S G D P T F I T T S -
 CTATGGCTGGTATGGGTAACGACGGTCCGTCACCGTGTACTAAGATGGACTACCGTTTCT
 3001 -----+-----+-----+-----+-----+-----+-----+ 3060
 GATACCGACCATACCCATTGCTGCCAGCAGTGGCCACAATGATTTCTACCTGATGGCAAAGA
 M A G M G N D G R H R V T K M D Y R F L -
 TGAACACTCTTGACAACATGGGTAACCTCACCAGAACCAAACTTGACAGTTCCTTTGGACTG
 3061 -----+-----+-----+-----+-----+-----+-----+ 3120
 ACTTGTGAGAACTGTTGTAGCCATTGAGTGGTCTTGGTTTGAACCTGTCAAGAAACCTGAC
 N T L D N I G N S P E P N L T V L W T D -

PCT/US05/27239 258/487

Figure 101G

```

3121 ~ACAAATGCCCATAACAACCTCCGTCCTACTGTATGCACATGAGCCACAACACTCTTCTA
-----+-----+-----+-----+-----+-----+-----+
3180 TGTTTAACGGTATGTTGAAGGCAGCCGATGACATACCGTGTACTCGGTGTTGTGAGAAGAT
-----+-----+-----+-----+-----+-----+-----+
c K L P Y N F R R Y C M H M S H K H S S I -
-----+-----+-----+-----+-----+-----+-----+
3181 TCCAATACGAAGGTGTAACAACAATGGCTAAAGACGGATATGGTGAAATGAGCTGTATCT
-----+-----+-----+-----+-----+-----+-----+
3240 AGGTTATGCTCCACATGTTGTTACCGAATTTCTGCCCTATACCACCTTTACTCGACATAGA
-----+-----+-----+-----+-----+-----+-----+
c Q Y E G V T T M A K D G Y G E M S C I S -
-----+-----+-----+-----+-----+-----+-----+
3241 CATGCTGTGTCTCCACTTGATCCAGAAATGAAGAACAACGCCACAACATCCAGTACT
-----+-----+-----+-----+-----+-----+-----+
3300 GTACGACACACAGAGGTGAACHTAGGCTTTTACTTCTTGTTCGGGTGTTGATGTCATGA
-----+-----+-----+-----+-----+-----+-----+
c C C V S P L D P E N E E Q R H N I Q Y F -
-----+-----+-----+-----+-----+-----+-----+
3301 TCGGTGCTCGTGTAAACGTTCTTAAAGCCCTTCTTACTGGTTTGAATGGTGGTTACGACG
-----+-----+-----+-----+-----+-----+-----+
3360 AGCCACGACACATTTGCAAGAATTTGGGAAGAATGACCAAACTTACCACCAATGCTGC
-----+-----+-----+-----+-----+-----+-----+
c G A R V N V L K A L L T G L N G G Y D D -
-----+-----+-----+-----+-----+-----+-----+
3361 ATGTTCAAAGACTACAAGTATTTGATATCGAACCAATCCGGTGACGGAAGTTCTTGAAT
-----+-----+-----+-----+-----+-----+-----+
3420 TACAAGTGTCTGATGTTTCATAAACHATATAGCTTGGTTAGGCACCTGCTTCAAGAACTTA
-----+-----+-----+-----+-----+-----+-----+
c V H K D Y K V F D I E P I R D E V L E F -
-----+-----+-----+-----+-----+-----+-----+
3421 TTGAATCAGTTAAAGCGAACTTTGAAAATCTCTTGACTGGTTGACTGACACTTACGTTAG
-----+-----+-----+-----+-----+-----+-----+
3480 AACTTAGTCAATTTCCGTTGAAACTTTTATAGAGAACTGACCAACTGACTGTGATGCATC
-----+-----+-----+-----+-----+-----+-----+
c E S V K A N F E K S L D W L T D T Y V D -
-----+-----+-----+-----+-----+-----+-----+
3481 ATGCCCTTGAACATCACCACATGACTGATAGGTACAACATACGAAGCTGTCAATGG
-----+-----+-----+-----+-----+-----+-----+
3540 TACGGAACTTGTAGTAGGTGATGACTGACTATCCATGTTGATGCTTCGACAAAGTTTACC
-----+-----+-----+-----+-----+-----+-----+

```


PCT/US05/27239 259/487

Figure 101H

C A L N I I H Y M T D R Y N Y E A V Q M A -
 3541 CCTTCTTGCCCAACTAAACAAGGTCGCCAACATGGGATTCGGTATCTGTGGATTGCTAACA
 -----+-----+-----+-----+-----+ 3600
 GGAAGAACGGTTGATTTGTTGCACGGTTGTACCCCTAAGCCATAGACACCTAAACCGATTGT

 C F L P T K Q R A N M G F G I C G F A N T -
 3601 CTGTTGATACATGTCAGCTATCAAAATACGCTACAGTTAAACCAATCCGGTGACGAAGATG
 -----+-----+-----+-----+-----+ 3660
 GACAACTATGTAACAGTCGATAGTTTATGCGATGTCATTTGGTTAGGCACCTGCTTCTAC

 C V D T L S A I K Y A T V K P I R D E D G -
 3661 GCTACATCTACGATTACGAAACAATCGGTGACTACCCACGCTGGGGTGAAGATGACCCAC
 -----+-----+-----+-----+-----+ 3720
 CGATGTAGATGCTAAATGCTTTGTTAGCCACTGATGGGTGGACCCCACTTCTACTGGGTG

 C Y I Y D Y E T I G D Y P R W G E D D P R -
 3721 GTTCAAACGAATGGCAGAAATGGTTGATCGAAGCTTACACAACCTCGTCTACGTAGCCACA
 -----+-----+-----+-----+-----+ 3780
 CAAGTTTGCCTTAACCGTCTTACCCTACTAGCTTCGAAATGTTGAGCAGATGCATCGGGTGT

 C S N E L A E W L I E A Y T T R L R S H K -
 3781 AACTATACAAGACCGCAGAAAGCTACAGTATCACCTTTTGACAATCACATCTAACGTTGCTT
 -----+-----+-----+-----+-----+ 3840
 TTGATATGTTTCTGCGTCTTCGATGTCATAGTGAAAACCTGTTAGTGTAGATTGCCAACGAA

 C L Y K D A E A T V S L L T I T S N V A Y -
 3841 ACTCTAAACAACCTGGTAACCTCACAGTTCACAAGGTGTATACCTCAACGAAGATGGTT
 -----+-----+-----+-----+-----+ 3900
 TGAGATTTGTTGACCATTTGAGTGGTCAAGTGTTCACACATATGGAGTTGCTTCTACCAA

 C S K Q T G N S P V H K G V Y L N E D G S -
 3901 CTGTGAACCTTGTCTAAACTTGAATTTCTCTCACAGGTGCTAACCCATCTAACAAAGCTA
 -----+-----+-----+-----+-----+ 3960

PCT/US05/27239 260/487

Figure 101I

```

GACACTTGAACACAGATTTGAACCTTAAGAAGAGTGGTCCACGATTGGGTAGATTTGTTCCGAT
V N L S K L E F S P G A N P S N K A K -
3961
AAGGTGGTTGGTTGCAAAAACHTGAACCTCACCCTTCTAGCCCTTGACTTTAGTTATGCAGCTG
-----+-----+-----+-----+-----+-----+-----+
TTCCACCACAACCGTTTTGAACTTGAAGTGAAGAATCGGAACTGAATAATACGTCGAC
-----+-----+-----+-----+-----+-----+-----+
4020
G G W L Q N L N S L S L D F S Y A A D -
4021
ACGGTATCTCATTGACTACACAAGTATCACCCTCGCGCTTTGGTAAGACTCGTGATGAAC
-----+-----+-----+-----+-----+-----+-----+
TGCCATAGAGTAACTGATGTTTCATAGTGGAGCGGAGAACCATTCCTGAGCACACTACTTG
-----+-----+-----+-----+-----+-----+-----+
4080
G I S L T T Q V S P R A L G K T R D E Q -
4081
AAGTTGATAAATTGGTAACAATCCTTGATGGTTACTTCGAAAACGGTGGACACACCGTTA
-----+-----+-----+-----+-----+-----+-----+
TTCAACTATTGAACCATTGTTAGGAACCTACCAATGAAGCTTTTGGCCACCCTGTTGTGCAAT
-----+-----+-----+-----+-----+-----+-----+
4140
V D N L V T I L D G Y F E N G G Q H V N -
4141
ACTTGAACGTTATGGACTTGAACGATGTTTACGAAAAAATCATGTACGGGAAAGACGTTA
-----+-----+-----+-----+-----+-----+-----+
TGAACTTGCAATACCCTGAACCTTGCCTACAATGCTTTTTTTTAGTACAGTCCGCTTCTGCAAT
-----+-----+-----+-----+-----+-----+-----+
4200
L N V M D L N D V Y E K I M S G E D V I -
4201
TCGTACGTATCTCTGGATACTGTGTAACACTAAATACCCTCACCAGAACAAAAAACTG
-----+-----+-----+-----+-----+-----+-----+
AGCATGCATAGAGACCCTATGACACATTTGTGATTTATGAGGTGAGGTCCTTTTTTTTGAC
-----+-----+-----+-----+-----+-----+-----+
4260
V R I S G Y C V N T K Y L T P E Q K T E -
4261
AATTGACACAACGTTCTTCCACGAAGTCTTCAATGGATGACGCCCTTGGATGCATTGA
-----+-----+-----+-----+-----+-----+-----+
TTAACTGTGTTGCACAGAAGGTTCTTCAAGAAAAGTTACCCTACTGCGGAACCTACGTTAACT
-----+-----+-----+-----+-----+-----+-----+
4320
L T Q R V F H E V L S M D D A L D A L S -

```

PCT/US05/27239

261/487

Figure 101J

4321 GCTAATCAAGTTCITGAATAAATAAGGGCTCTTTGTCAACTGTAGTGGTGAAGAAA + 4380
 -----+-----+-----+-----+-----+-----+-----+
 CGATTAGTCAAGAACTTATTTATTTCCCGAGAAACAGTTGACATCACCCCAACTTCTTT

c *
 4381 AGCTAAGCTCGAGAAAGGACAAATTTTGCCCTTTCTTTTGAATGTTTCAGAGCGGATGAAA + 4440
 -----+-----+-----+-----+-----+-----+-----+
 TCGATTGAGCTCTTCCCTGTTTAAACAGGAAAGAAAAACTACAAAGTCTCGCTACTTT

e * A R S L V F K T R E K K I N L A I F -orf1_670 homologue of sp0460, transposase
 4441 ATCCGTTTTTTGAAAGTTTCAAAGTTCGGAAAAACCAAGGCAATGCGCTTGATGCTTTTG + 4500
 -----+-----+-----+-----+-----+-----+-----+
 TAGGCAAAAACCTCAAAGTTTCAAGGCTTTTGGTTTCCGTAACGGGAACACTACAGAAAC

e I R K K F N E F N R F G F A N R K I D K -
 4501 ATGAGTTTGTAGTGGCCCTCAAGTTTAGCGTTAGAATAAGGCAATTC AATGGCGTTAGTG + 4560
 -----+-----+-----+-----+-----+-----+-----+
 TACTCAACAATCACCGGAGTTC A AATCGCAATCTTATTCGGTTAAGTTACCGCAATCAC

e I L K N T A E L K A N S Y P L E I A N T -
 4561 ATGTAGTTTTATAGCAATAAATGTGCTCAAAGTGGTTTAAAGGTCCGGTTGAGATGA + 4620
 -----+-----+-----+-----+-----+-----+-----+
 TACATCAAAAATAATCGTTTATTTACACGGATTCACCAAAAATTTCCACGCCAACTCTACT

e I Y N K Y C I F T S L T T K F T R N L H -
 4621 GGTAACGTGCTTGAATAAAGCCCAAACTGGTCAGTATTTCTCTCTGTAGATGAAAT + 4680
 -----+-----+-----+-----+-----+-----+-----+
 CCATTGCCACAGAACTTAATTCGGGTTTGGACCAATAGAGAGAGAACATCTACTTTA

e P L T D Q I L G W F Q D T N K E Q L H F -
 4681 AGGAGTAGTTGATACAGGTCATAGTAATCTTTAAGTTTCAGGTAAGTAAGATTTTC + 4740
 -----+-----+-----+-----+-----+-----+-----+
 TCCTCATCAACTATGTCAGTATCATTAGAAAATTC AAGTCCATGATCTCATTTCTAATAAG

PCT/US05/27239

262/487

Figure 101K

```

e L L L Q Y L D Y Y D K L E P V L T F I K -
  TTCAGACACTCCCTAGGAGTTAAGGTCCTCTGAAAGTTCTAGCATAGAAAGGCTTAAGA + 4800
  AAGTCTGTGAGGGATCCCTCAATCCAGAGAGACTTTCAGATCGTATCTTTCCGAAATCT
e K L C E R P T L T E R F T R A Y F P K L -
  GAGAGTTCCCGACTATCTTTAGGATAAATTCAGTAATTTAAGAGCTCTGTATTC + 4860
  CTCCTCAAAGCGTGATAGAAAATCCCTATTTAAAGGTCATTTATAAATTCGAGACATAAGG
e S L K R S D K L I F K W Y Y K L A R Y E -
  AGAGATTTATCAATAAATGCTTCATGATGTTGATTTCTAGTCTGATTAAGAGCCCTGCTC + 4920
  TCCTCAAATAGTAGTTTAAACGAAGTACTACAATAAGATCAGACTAATTCCTCGGACGAG
e L S K D D F Q K M I N I R T Q N L A R S -
  ATGTGTTGGACAATGTGGAAACGATCGAGAACAATTTTAGCATTTGGGAAATAATTTCTTA + 4980
  TACACAACCTGTTACACCTTTGCTAGCTCTTGTAAATAACGTAACCCCTTTATTAAGAAT
e M H Q V I H F R D L V I K A N P F L K K -
  ATGAGAGGATATAACTTCCAGACATATCAACAGTGACGACTTTAACTTTTCTTAGCT + 5040
  TACTCTCCCTATATTGAAGGTCGTGATAGTTGTCACCTGCTGAAATTTGAAAAAAGATCGA
e I L P I Y S G S M D V T V K V K R A -
  TCTTTCCGAGTACTTGAAGAAATGATTTCCGGATGGTTGTTGACGCTGTATCAAGAATG + 5100
  AGAAAGCTCATGAACTTCTTTACTAAAGCCCTACCAACAACCTGCAGACATAAGTCTTAC
e E K S Y K F F H N R I T T Q R R N D L I -
  GTCATGATTTTCTTAGTGTGAAATCTGAGCAATGAAAGCCATTTCCCTTCTGGTAG + 5160

```

Figure 101L

CAGTACTAAAGAATCACAACCTTAGGACTCGTTACTTTCGGTTAAAGGGAAGACCCTC
 e T M I K K T N F D Q A I F A L K G K Q Y -
 GAGAAATTCATCCAGGAGAGGATTCAGGCAAGTGGTGAATCCTCTTGAAATGAAAT
 5161 -----+-----+-----+-----+-----+ 5220
 CTCTTAAGTAGGGTCTCTCTCCTAAAGTCCGTTTCACCACATAGGAGAACCCTTACTTTA
 e S F E D W S L I E P L T T Y D E Q F H F -
 TGCTTGAGCTTACGATAGACGGTAGAGGTAGAGATGGCTAATTTAGAAGCGATA
 5221 -----+-----+-----+-----+-----+ 5280
 ACGAACTCGAATGCTATCTGCCATCTCCATCTCCATCTCCGATTAATAATCTTCGCTAT
 e Q K L K R Y V T S T S I A L K S A I -
 TGTGTAAGAGCCCTCTCTGTTGAGTAGGAGTTGGGCAATTTCTGTCTCACCATTTCCGAG
 5281 -----+-----+-----+-----+-----+ 5340
 ACACATCTCGGAGAGACAACCTCATCTCAACCCGTTAAAGACAGAGTGGTAAAGGCTC
 e H T L A E R N L L Q A I K Q R V M E S -
 ATTTGGCAATTTCTGAACGAGAGTTGTTTCAGCTACAGTGACTTTCGACAGGACTTG
 5341 -----+-----+-----+-----+-----+ 5400
 TAAACCGTTAAAAGACTTGGCTCTCAACAAGTCCGATGCTACTGAAAGGCTGCCGTAAC
 e I Q C N K Q V L T F E A V T V K R C S K -
 CATTGAAAATCGTCTCTTTTCAAAATGAATGAGGCTAGGGAACCCACCAATCTCGATAAAA
 5401 -----+-----+-----+-----+-----+ 5460
 GTAACCTTAGCAGAGAAAAGTTTACTTACTCCGATCCCTTTGGTGGTTAGAGCTATTTT
 e C Q F R R K K L H I L S P F G G I E I F -
 GGGATTTTGAAGGCTTTTGGAAAGTCGTAATTTGATTTTCCCTTTACAGTGTTTACAT
 5461 -----+-----+-----+-----+-----+ 5520
 CCTAAAATCTTCCGAAAACCTTCAGCATAAACAATAAAGGAAATGTCACAAAATGTA
 e P I K S P K Q F D Y K I Q K G K C H K C -

PCT/US05/27239 264/487

Figure 101M

5521 TTAGGTGGTGATAATCAAGTGTAGCGAAGACTTCGATATGGTATCGTGTGAATGGCT + 5580
 -----+-----+-----+-----+-----+-----+-----+-----+
 5581 AATCCACCACATATAGTTCACATCGCTTCTGAAGCTATATACCCATAGCAGACTTACCGA

e K P P H Y D L T A F V E I H T D H Q I A -
 TTATTTAAGGTGATGTTTTTGTCTTTTATTCCGATGAGTAATGTGGTATGATGATGTTGT
 -----+-----+-----+-----+-----+-----+-----+
 5640 AATAAATCCACTACAAAACAGAAAATAAGGCTACTCATACACCATACTAATACTACACA

e K N L T I N K D K I G I L L T T H N I H -
 TCCATAAGATACTTTCTAATGAGTTGTTTAGGGCTTTTCATFATAAAGTCTTATGGGACT + 5700
 -----+-----+-----+-----+-----+-----+-----+
 5641 AGGTATCTATGAAAGATTACTCAACAAATCCGGGAAAAGTAATATTCAGAATACCCCTGA

e E M -
 TTTTGTGATACTCAAAAAGCCCTATAATCTCCACAGTGGGATTTACCCACTACAGAAATTA + 5760
 -----+-----+-----+-----+-----+-----+-----+
 5701 AAAAACTATGAGTTTTTCCGGGATATTAGAGGGTGCACCCCTAAAATGGGTGATGCTTTTAT

TAGAGCCAGAAAACACTTTTGTTCACCTAGCAGAACTAGAGAGCAGAAAGTGTTTTCT + 5820
 -----+-----+-----+-----+-----+-----+-----+
 5761 ATCTCGGCTTTTTTGTGAAAACAAGTATCGTCTTTGATCTCTCGTCTTCACAAAAGA

5821 GTTCAGATTTACCCAAAACITGGGAAATATGGGGATAAAGAATAGAGATGGCTTAGGAAGCC + 5880
 -----+-----+-----+-----+-----+-----+-----+
 5880 CAAGTCTAAATGGGTTTTGACCCCTTATACCCCTATCTTTATCTCTACCCGAATCCTTCGG

5881 CCTTTTGTGTGTAGACAGTACGATGAACCTTATAACAATAAGTGTAGCCTTTGTAGCAATC + 5940
 -----+-----+-----+-----+-----+-----+-----+
 GGAAAAACACACAGATCTGTCACTGCTACTTGAATATTGTTTATCACTCGGAAAATACTGTTAG

e ATTGGGACCCGTTTTGTCAAAAAGCCCTTTTTCCGGATATCTACAATGTCTGATAGATGAGA
 * L L Y H A K K A I -orf2_670 homologue of sp0461, transcriptional regulator

PCT/US05/27239

265/487

Figure 101N

```

5941 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TAACGCTGGCAACACAGTTTTTCGGAGAAAAGCCCTATAGATGTTAACAGACTATCTACTCT
      M A V R K D F A E K R I D V I T Q Y I L -
      CGCTGTGGCTAACATGCAAACTAAGCAATCGTCAAAAAGTATGTTCCCTTTGGGA
6001 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      GCGACAACCGATTGTACGTTTAGATTCCTGTTAGCAGTTTTTCACTACAAGGGAACCCCT
      R Q Q S V H L D L A I T L F H H K G K P -
      TACTGCTTTTAAACGTAAGCAGGTAATCTTTCGTTGTAATAATAATAATCAATGGCTCTGTC
6061 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      ATGACGAAAATAATGCAATCCGTCATAAAGAAAGCAACATTAATATTAGTTACCGAGACAG
      Y Q K K V Y P L Y E K T T I I I L P E T -
      AAATGCTCCTCTGAAGGAGGAGGACTAATTAGATAATTTGTAATCCTGTAAACAGAGGCAACT
6121 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TTTACGAGGAGACTTCCCTCCCTGATTAATCTTATAACATAGGACATTTGTCTCCGTTGA
      L H E E S P P P S I L I N Y G T V S A V -
      TTGTCAGTAAAATTCCGTAATAAATGGACTTTTATTAAGTTTACATCTGCTTGATTTATTT
6181 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      AACAGTCATTTTAAGGCATTTTATTACCTGAAATAATTCAAATGTAGACGAACATAATAA
      K D T F N R L I I S K I L N V D A Q N N -
      AAAATGATAAAAATCGGGATAGCAGGTAGTGAGGAAAAGATGGTTTCTGTCAAGTAGAGT
6241 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TTTTACTATTTTAGCCCTATGGTCCCATCACTCCCTTTTCTACCAAGACAGTTCACTCA
      L I I F I P I A P L S S F I T E T L Y L -
      GAGAAAAGGTACAGCCGATGCTGGTCGATAACTCCTTCAATCTTCTGCTCAGTCATCCAC
6301 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      CTCCTTTCCATGTCGGCTACGACCAGCTATTGAGGAAGTTAGAACGAGTCAGTAGGTG
      S F L Y L R H Q D I V G E I K Q E T M W -

```

266/487

Figure 1010

6361 TCTTGAACAATTCCTTTCGAAATATGATACAGTGGCTTGTGCGTTTCAATCCCATAAATGT 6420
-----+-----+-----+-----+-----+-----+-----+
AGAACTTGTTAACGAAAGCTTTATACTATGTCCACGACACGCGAAAGTTAGGGTATTACA
e E Q V I A K S I H Y L P K D S E I G Y H -
6421 TCGTAATAATATAATAGGGAACAGATTTGTAAACCAAAACGTTCTTGTAAAG 6480
-----+-----+-----+-----+-----+-----+-----+
AGCATTATTAATAATATCCCTTGAATCTAAACATTTGGTTTGTGCAAGAACAATTC
e E Y Y N Y Y P V L N Q L G F L F T R T L -
6481 AAAAGTCAGTGGTAAAAAAGAAAGAGAAATCGAAATGTCATTTCCSTAAGATATCTTG 6540
-----+-----+-----+-----+-----+-----+-----+
TTTCAGTCACGACAATTTTTTCTTTCTTAAGCTTTACAGTTAAGGATTCATAAAGAAC
e F T L A T L F S L S N S I D N G L I N K -
6541 AACTTGGATAGTAGATGCTTTCCTTGTATGCTGAAGAATCAGTTGAATAGTATGAGTC 6600
-----+-----+-----+-----+-----+-----+-----+
TTGAACCTATCATCTACGAAAGGAGAACATACGACTTCTTAGTCAACTTATCATACTCAG
e F K S L L H K G R T H Q L I L Q I T H T -
6601 TTTTTCCTTGATTCCTTGTCTTGGAAAACGAAAGATTAGCAGAACAAATAACCAA 6660
-----+-----+-----+-----+-----+-----+-----+
AAAAAAGAACTAAGGTAAACAGGAACCTTTTCTTCTTAATCGTCTTGTATTGTTT
e K K E Q N W K D K S F S S N A S C Y V L -
6661 AAGATATAATCCAGTTCCTCCAGTAAAGTCAATGTTGGCATGTTGGCTCTAAGTAAGTT 6720
-----+-----+-----+-----+-----+-----+-----+
TTCTATATTAGGTCAGAAGGACTCATTTTCAGTACACCCGTACACCCGAGATTCATTCAA
e F I Y D L E E Q T F T M N A H P E L Y T -
6721 TGGCAATGTTCCATCAAAAATCGGATACATAAAGAGGTTTTTTAATTTTCAAACTCTTTG 6780
-----+-----+-----+-----+-----+-----+-----+
ACCGTTACAAGGTAGTTTTAGCCATGATTTCTCCAAAAAATTAATAAAGTTTGAGAAAC

PCT/US05/27239

267/487

Figure 101P

e Q C H E M L I P Y M F L N K L K E F E K -
 GACTCAGGGAACCTCAAGTGGAAATCCCGACGTTTCCAAAGTGAGTGCCACTAGTATGCTA 6840
 -----+-----+-----+-----+-----+
 CTGAGTCCCTTGAGTTCACCTTTAAGGCTGCAAGGTTCACTCACCGTGATCATACGAT
 e S E P F E L P F E R R K W T L A V L I S -
 AAATGAACATACTCGTCAGGTGGATTTCTAAACAGTTCATGACTGAGTTGAGAATTAGAC 6900
 -----+-----+-----+-----+-----+
 TTTACTTTGATGAGCAGTCCACACTAAAGATGTCAAGTACTGACTCAACTCTTAATCTG
 e F H V Y E D P T I E L L E H S L Q S N S -
 TGCACAATCATATGTTGACCCCAATCCATACTTCCATCATTAATAATCAATAAATCTCAATA 6960
 -----+-----+-----+-----+-----+
 ACGTGTAGTATACACACTGGTGGTGGTATGAAGGTAGTAAGTTTAGTATTAGAGTTAT
 e Q V I M H T V W D M S G D N L D Y I E I -
 CCAAAATGAAACTGGAGGAGTCAATTAATAAACCAGTATCCGATATTCAGGACCAACTACT 7020
 -----+-----+-----+-----+-----+
 GGTTTACTTTGACCTCCCTCAGGTTAATTTTGTGCTTACGCTATAAGTCCCTGGTTGATGA
 e G F H F Q L L A I L F R I R Y E P G V V -
 TGATTTTTCACAAGGTTCCAAACCTACTGAACGTTAGTAACAAGCCACACTTTTGTCTGACG 7080
 -----+-----+-----+-----+-----+
 ACTAAAAAGTGTCCAGGTTTGGATGACTTGCATCATTTGTTCCGGTGTGAAAACAGCATGC
 e Q N K V L D L G V S R L L L G C K Q R V -
 CGGTAGCCCTGTTGCGATGGAAATATACTCTTTTGTGTAAATTCGTTAAAGCTTTGATTA 7140
 -----+-----+-----+-----+-----+
 GCCATCGGACAAACGCTACCTTTATATGAGAAAAACACATTTAAGCAATTTGGAACATAAT
 e R Y G T A I S I Y E K Q T F E N F S Q N -
 CCTTGTAGTAAAGAGCGGAGTATTTTAAAAATAGTTGATTGGTTATAAAGCTGATGG 7200
 -----+-----+-----+-----+-----+

PCT/US05/27239

268/487

Figure 101Q

GGAAACATCATCTTCTCGCCCTCATAAAATTTTATCAACCTAACCAATATTCGACTACC
e G Q L L F F R L I K L I T S Q N Y L Q H -
7201 AAGTAATAATTCGTTGATGAGAATGGTTCGATTAATGAACCTTGTTCCTATCTAAA
-----+-----+-----+-----+-----+-----+ 7260
TTCAATTATAAGCAAACTACTCTTTACCACAAAGCTAATTAACCTTGAACAACGCATAGATTT
F Y Y N T Q H S H H E I L Q V Q Q T D L -
7261 TTAAAATGTCAACTCTTCCCGAATGTTCTTGTAAATCCCTGCAAAATGCTTAGGAGACTTT
-----+-----+-----+-----+-----+-----+ 7320
AATTTACAGTTGAGAAGGAGCTTACAAGAACAATTAAAGGACGTTTTTACGAATCCTCTGAA
N F T L E E E F T E Q L E Q L I S L L S -
7321 TTAGATTGTAATGAAGTTAAAAGTAGACAGTTCATCTAGTTCAAATAGACCGAATATCCAAAT
-----+-----+-----+-----+-----+-----+ 7380
AATCTAACATTTACTTCAATTTTCATCTGTCAAGTAGATCAAGTTATCTGGCTTATAGGTTA
K S Q L S T L T S L E D L E I S R I D L -
7381 AATAATTTAAAATGGTAATTTTATCTGTAAATCTTTTTTCAATGATTTTGTTCAGCATA
-----+-----+-----+-----+-----+-----+ 7440
TTATATAAAATTTTACCATTAAAATAGACATTAAGAAAAGTTACATAAACAAATCGTAT
L I N L I T I K D T I R K E I Y K N L M -
7441 GTTACCGAATCTTAGTTGCATATAGATAATTTTAAATTTATTAATAACAAGAAACTAAT
-----+-----+-----+-----+-----+-----+ 7500
CAATGGCTTAGAATCAACGTTATCTATTAATAAAATTAATAATATTTCTTTCTTTGATTA
TGTCCTGTCAAAAAGGTTGTGGAATTTCCGACTTTTATGATAAAACAGCATGTAATAAAA
7501 -----+-----+-----+-----+-----+-----+ 7560
ACAGAACAGTTTTTCCCAACACCTTAAAGGCTGAAATAACTATTTTGTGTCATATATTTT
GGCATTTTAAAGATAGTAATGAGTATTTGGTGGAGTTTTTATGGCTTATTTTTTTTATTAGA
7561 -----+-----+-----+-----+-----+-----+ 7620

PCT/US05/27239 269/487

Figure 101R

CCGTAAATTTCTATGATTAATCAACCCACCTCAAAATACCGAATAAAAAATAATCT
 7621
 AAATATTTTTTATCAAAATATGTCGTTCTATAAAAAAATATGTGATAAAAATATCTATT
 TTTATATAAAAAATAGTTTTTATAACAGCAAGATATTTTTTATACACTATTTTATAGATAA
 7680
 GTGATGGAAGTTGTTTTAATTTAATACTAGGATAGTTAATAGTAATACTATACCTAT
 7740
 CACTACCTTCAACAAAATTAATAATGATCCCTATCAATTAATCATATGATATGATATGATA
 ATGTATACAAAGTGTGTCATGCCAGGTTGAGAAGATAGCTATAACGCACCTTTTATACGC
 7741
 TAACATATGTTCAACACAGTAACGGTCCAACTTCTATCGATATGCGGTGAATAATGCG
 TTTTGTACGTTTGTAGTGAACGGATTAACCTCAGTGAGATAAATTTTATCAGAACAATA
 7801
 AAAACGATGCAAAACAATCACCTTGCCTAATGAGTCACTCTATTTAAATAATGCTTGTATT
 GTAATCCGTTTCTTCGTGTATACAGATTGAAAAGTACCTATGAATCATAGAAAGGATTAAC
 7861
 CATTAGCCAAAGAAGCACATATGTCCTAACTTTTCATGGATACCTTAGTATCTTCCTAATTGA
 TGTCTATGAATAATGCTTAAACAGGGAGACACACATGAAAAAAGTAAGAAAGATATTCA
 7921
 ACAAGATACCTTATACGAATTGTCCTCCCTGTGTGTTACTTTTTTTCATTTCTTCTATAAAGT
 M L N R E T H M K K V R K I F Q -orf3_670 homologue of sp0462, LPXTG
 GAAGGCAGTTGCAGGACTGTGCTGTATATCTCAGTTGACAGCTTTTTTCTTCGATAGTTGC
 7981
 CTTCCGTCAACGCTGACACGACATATAGAGTCAACTGTCCGAAAAAGAACGATATCAACG
 K A V A G L C C I S Q L T A F S S I V A -

270/487

Figure 101S

8041 TTTAGCAGAAACCCCTGAAACCAGTCCAGCGATAGGAAAAGTAGTATTAAGGAGACAGG + 8100
 -----+-----+-----+-----+-----+-----+-----+
 AAATCGTCTTTGGCGACTTTGGTCAGGTCGGTATCCTTTTTCATCACAATAATTCCTCTGTGCC
 b L A E T P E T S P A I G K V I K E T G -

8101 CGAAGGAGGAGCGCTTCTAGGAGATCCCGTCTTTGAGTTGAAAAACAATAACGGATGGCAC + 8160
 -----+-----+-----+-----+-----+-----+-----+
 GCTTCCCTCCGGAAGATCCCTACGGCAGAAACTCAACTTTTGTGTTATGCCCTACCCGTTG
 b E G G A L L G D A V F E L K N N T D G T -

8161 AACTGTTTCGCAAGGACAGGCGCAACAGGAGAACCGGATATTTCAAAACATAAAACC + 8220
 -----+-----+-----+-----+-----+-----+-----+
 TTGACAAAAGCGTTTCCCTGCTCCCGGTTTGTCCCTATATAAAAGTTTGTATTTTGG
 b T V S Q R T E A Q T G E A I F S N I K P -

8221 TGGACATACACCTTGACAGAAAGCCCAACCTCCAGTTGGTTATAAACCCCTCTACTAAACA + 8280
 -----+-----+-----+-----+-----+-----+-----+
 ACCCTGTATGTGGAACTGCTTCCGGTTGGAGGTCAACCAATATTTGGGAGATGATTTGT
 b G T Y T L T E A Q P P V G Y K P S T K Q -

8281 ATGGACTGTTGAAGTTGAGAAGAATGGTCGGACGACTGTCCAAGGTGAACAGGTAGAAAA + 8340
 -----+-----+-----+-----+-----+-----+-----+
 TACCTGACAACCTTCAACTCTTCTTACCAGCCCTGCTGACAGGTTCCACTTGTCCACTTTT
 b W T V E V E K N G R T T V Q G E Q V E N -

8341 TCGAGAAGAGGCTCTATCTGACCAGTATCCACAACAGGGACTTATCCAGATGTTCAAAC + 8400
 -----+-----+-----+-----+-----+-----+-----+
 AGCTCTTCTCCGAGATAGACTGGTCAATAGGTTTGTCCCTGAATAGGTTCTACAAGTTTG
 b R E E A L S D Q Y P Q T G T Y P D V Q T -

8401 ACCTTATCAGATTATTAAGGTAGTGGTTCGGAAAAAACAACGGACAGCACAAAGCGTTGAA + 8460
 -----+-----+-----+-----+-----+-----+-----+
 TGGAAATAGTCTAATAATTCCATCTACCAAGCCCTTTTGTCCCTGTTGTTCCCGCAACTT

PCT/US05/27239

271/487

Figure 101T

b P Y Q I I K V D G S E K N G Q H K A L N -
8461 TCCGAATCCATATGAACGGTGTGATCCAGAAGGTACACATTTCAAAGAGAATTTATCAAGT + 8520
AGGCTTAGGTACTTGCACACTAAGGCTTCCATGTGAAGTTTCTTAAATAGTTCA
b P N P Y E R V I P E G T L S K R I Y Q V -
8521 GAATAATTGGATGATAACCAATAATGGAATCGAGTTGACGGTTAGTGGTAAACGACGGT + 8580
CTTATTAAACCTACTATTGGTTATACCTTAGCTCAACTGCCAATCACCAATTTTGCTGCCA
b N N L D D N Q Y G I E L T V S G K T T V -
8581 TGAAACGAAAGAAGCCCTCTACTCCGCTAGATGTTGTTATTCTATTAGATAACTCCAAATAG
ACTTTGCTTCTTCCGGAGATGAGCGGATCTACAACAATAAGATAATCTATTGAGGTTATC
b E T K E A S T P L D V V I L L D N S N S -
8641 TATGAGTAATATTCGACATAATCATGCCATCGAGCGGAAAAGCGGAGAGCCGACACCG + 8700
ATACTCAATTATAAGCTGTATTAGTACGGGTAGCTCGCCCTTTTTCGCCCTCTTCGGCTGTGC
b M S N I R H N H A H R A E K A G E A T R -
8701 AGCCCTTGTAGATAAGATTACCTCCAATCCAGATAATCGAGTAGCACCTTGTGACTTATGG + 8760
TCGGGAACATCTATTCTAATGGAGGTTAGGTCATTATTAGCTCATCGTGAACACACTGAATACC
b A L V D K I T S N P D N R V A L V T Y G -
8761 CTCAACTATCTTTGACGGTTCAGAAGCTACTGTGGAAAAGGGGTAGCAGATGCGGAACGG + 8820
GAGTTGATAGAAAACCTGCCAAGTCTTCGATGACACCTTTTTCCTCCCATCGTCTACGCTTGCC
b S T I F D G S E A T V E K G V A D A N G -
8821 AAAAAATATTGAATGACTCAGCTTTATGGACGTTCCGATCGTACGACGTTTACAGCTAAAC + 8880

Figure 101U

TTTTATAACTTACTGAGTCGAAATACCTGCAAGCTAGCATGCTGCAAAATGTCGATTTTG
 K I L N D S A L W T F D R T T F T A K T -
 8881
 TTATAATTATAGCTTTTAAATCTCACATCAGATCCTACTGATATTCAAAATTAAGGA
 AATATTAATATCGAAAATTTAGAGTCTAGTCTAGGATGACTATAAGTTTGATAATCCCT
 Y N Y S F L N L T S D P T D I Q T I K D -
 8941
 TAGGATTCACATCAGATGCAGAGGAATGAACAAGACAAAATGATGATCAATTCGGCGC
 ATCCTAAGGTAGTCTACGTCCTTAACTTGTTCCTGTTAACTACATAGTTAAGCCGCG
 R I P S D A E E L N K D K L M Y Q F G A -
 9001
 GACTTTTACCCAGAGGCTTTGATGACCGCTGATGATATCTTGACAAGCAGGCAAGACC
 CTGAAAATGGGTCCTCCGAAACTACTGGCGACTACTATAGAACTGTTTCGGTCCGTTCTGG
 T F T Q K A L M T A D D I L T K Q A R P -
 9061
 AAACAGTAAAAGGTTATTTCCACATACAGATGGTGTCCGACTATGTCATATCCAAAT
 TTTGTCAATTTTCCAAATAAAGGTGTAATGTTTACCACAAAGGCTGATACAGATAGGTTA
 N S K K V I F H I T D G V P T M S Y P I -
 9121
 TAATTTTAAATATACAGGAACGCGCAATCGTACAGAACTCAGCTGAATAATTTTAAAGC
 ATTAATAATTTATATGTCCTTGGTGGTTAGCATGCTTGTGAGTCCGACTTATTAATAATTTCCG
 N F K Y T G T T Q S Y R T Q L N N F K A -
 9181
 AAAAATCCAAATAGTAGCGGGATATTTACTGGAGGACTTTGTTACATGGTCAGCAGATGG
 TTTTGTGAGGTTTATCATCGCCCTATAATGACCTCCCTGAAACAATGTACCAGTCCGTCCTACC
 K T P N S S G I L L E D F V T W S A D G -

PCT/US05/27239

273/487

Figure 101V

9241 TGAACATAAGATTGTTCTGGAGATGGTGAAGTTATCAGATGTTTACGAAGAAACCTGT + 9300
ACTTGTATTCTAACAAGCACCTCTACCACTTTCAATAGTCTACAAAATGCTTCTTTGGACA
b E H K I V R G D G E S Y Q M F T K K P V -
9301 AACAGACCAATAACGGAGTTCATCAATACTTTCAATCACCTCCATGGAGCAGAGCTAA + 9360
TTGTCTGGTTATGCCCTCAAGTAGTTTATGAAAAGTTAGTGGAGGTACCTCGTCTCTCGATT
b T D Q Y G V H Q I L S I T S M E Q R A K -
9361 ATTAGTTTCAGCGGGAATATAGGTTCTATGGAACCTGACTTGTATTTATATTTGGCGTGATAG + 9420
TAATCAAAGTCCCTATATCCAAAGATACCTTGACTGAACATAAATAAACCACCTATC
b L V S A G Y R F Y G T D L Y L Y W R D S -
9421 TATTCAGCCATCCATTTAATCTAGTACCGATTGGATTACCAACCATGGTGACCCCTAC + 9480
ATAAGATCGGATAGGTAATGAGATCATGGCTAACCTAATGGTTGGTACCACCTGGGATG
b I L A Y P F N S S T D W I T N H G D P T -
9481 GACTTGGTATTATAACGGAAATATGGCTCAGGATGGCTATGATGCTTCTACTGTTGGGGT + 9540
CTGAACCATATAATGCTTATACCGGAGTCCCTACCGATACCTACAGAAAGTGACAACCCCA
b T W Y Y N G N M A Q D G Y D V F T V G V -
9541 TGGTGTAAACGGGGATCCCTGGTACGGATGAAGCAACGGCTACTAGATTATTCAGAGCAT + 9600
ACCACATTTGCCCTTAGGACCATGCCCTACTTCTGTTGCCGATGATCTAAATACGTCCTCGTA
b G V N G D P G T D E A T A T R F M Q S I -
9601 CTCCTAGTCTCCGACAACTACACTAAGTAGCAGATCCATCTCAGATTTTACAAGAATT + 9660
GAGATCAAGAGGACTGTTGATGATGCAATGCCATCGTCTAGGTAGATCTAAAATGTTCTTAA

PCT/US05/27239

Figure 101W

b S S S P D N Y T N V A D P S Q I L Q E L -
 9661 GAATCGCTACTTCTATATACTGTCATGAGAGAAATCTATCGAAAAATGGTACGATTAC + 9720
 CTTAGCGATGAAGATATGATAGCAGTACTCTTCTTTAGATAGCTTTTACCATGCTAATG
 b N R Y F Y T I V N E K K S I E N G T I T -
 9721 AGACCCGATGGTGAACATAATTGATTTCCAATTTGGGAGCAGATGGAAGGTTTGATCCAGC + 9780
 TCTGGCTACCCACTTGATTAACATAAAGGTTAACCCCTCGTCTACCTTCCAAACTAGGTCG
 b D P M G E L I D F Q L G A D G R F D P A -
 9781 GGATACACTTTAAC TGCAACGATGGTAGTTCGTTGGTGAATAATGTCCCTACTGGGGG + 9840
 CCTAATGTGAAATGACGTTTGCTACCATCAAGCAACCACCTTATTACAGGGATGACCCCC
 b D Y T L T A N D G S S L V N N V P T G G -
 9841 ACCACAAAATGATGGTGGCTTGTAAATAAAGTCAAAAAGTGTCTATGATACGACTGAGAA + 9900
 TGGTGTTTTACTACCACCGAACGATTTTACGTTTTCACAAGATACTATGCTGACTCTT
 b P Q N D G G L L K N A K V F Y D T T E K -
 9901 AAGGATTCGTTGTAACAGGTTTGTACCTTGGAACGGGTGAAAAAGTTACATTGACTTATAA + 9960
 TTCCTAAGCACATTTGCCAAACATGGAACCTTGGCCACCTTTTCAATGTAACGAAATTT
 b R I R V T G L Y L G T G E K V T L T Y N -
 9961 TGTTCGCTTGAATGACCAATTTGTAAGCAATAAATCTATGACACGAAATGGTCCGAACAAC + 10020
 ACAAGCGRACTTACTGGTTAAACATTCGTTATTAAAGATACTGTGCTTACCAGCTTGTGTTG
 b V R L N D Q F V S N K F Y D T N G R T T -
 10021 CCTACACCCTAAGGAAGTAGAAAGAACACAGTGGCGGACTTCCCGATTCCCTAAGATTCCG + 10080

Figure 101X

GGATGTGGGATTCCTTCATCTTTCTTGFGTCACGGCTGAAGGGCTAAGGATCTAAGC
L H P K E V E K N T V R D F P I P K I R -
TGATGTACGAAAGTATCCAGAAAATCACAATTCCAAAGAGAAAAAATCTGGTGAATGGA
10081 -----+-----+-----+-----+-----+-----+-----+-----+ 10140
ACTACATGCTTTCATAGGTCCTTTAGTGTAAAGGTTTCTCTTTTGAACCACTTTAACT
D V R K Y P E I T I P K E K L G E I E -
GTTTATTAAAGATCAATAAGAATGATAAAAAACCACCTGAGAGATGGGCTTTAGTCTTCA
10141 -----+-----+-----+-----+-----+-----+-----+ 10200
CAAAATAATTCTAGTTACTTACTATAATTTTGGTGACTCTCTACGCCAGAAAATCAGAAAGT
F I K I N K N D K K P L R D A V F S L Q -
AAAACAACATCCGGATATCCAGATATTATGAGCTATTGATCAAAAATGGCACATTATCA
10201 -----+-----+-----+-----+-----+-----+-----+ 10260
TTTTGTTGAGCCCTAATAGGCTATAAATACCCTCGATAACTAGTTTACCCTGGAATAGT
K Q H P D Y P D I Y G A I D Q N G T Y Q -
AAATGTGAGAACAGGTGAAGATGGTAAGTTGACCTTTAAAATCTGTCAGATGGGAAATA
10261 -----+-----+-----+-----+-----+-----+-----+ 10320
TTTACACTCTTGTCCACTTCTACCATTCAACTGGAAATTTTGTAGACAGTCTACCCCTTTAT
N V R T G E D G K L T F K N L S D G K Y -
TCGATTATTTGAAAAATTCCTGAACCAGCTGGTTATATAACCCGTTCAAATAAAGCCCTATCGT
10321 -----+-----+-----+-----+-----+-----+-----+ 10380
AGCTAATAAACCTTTAAGACTTGGTCGACCAATAATTTGGGCAAGTTTATTTCCGGATAGCA
R L F E N S E P A G Y K P V Q N K P I V -
TGCCTTCCAAAATAGTAAATGGAGAAGTCAGAGATGTGACTTCAATCGTTCACCAAGATAT
10381 -----+-----+-----+-----+-----+-----+-----+ 10440
ACGGAAGGTTTATCATTTACCCTTCTCAGTCTACACTGAAGTAGCAAGGTTGTTCTATA
A F Q I V N G E V R D V T S I V P Q D I -

Figure 101Y

10441 ACCAGCGGGTTACGAGTTTACGAATGATAAGCAGCTATATACAAATGAGCCAAATCCCTCC
 -----+-----+-----+-----+-----+-----+-----+
 10500 TGGTCGCCCAATGCTCAAATGCTTACTATTCGTGATATAGTGTTFACCTCGGTTAAGGAGG

b P A G Y E F T N D K H Y I T N E P I P -

10501 AAAAAAGAGAAATATCCCGAACTGGTGGTATCCGGAATGTTGCCATTCTATCTGATAGGTTG
 -----+-----+-----+-----+-----+-----+-----+
 10560 TTTTTCCTTATAGGAGCTTGACCACCAATAGCCCTTACAACGGTAAAGATAGACTATCCCAAC

b K R E Y P R T G I G M L P F Y L I G C -

10561 CATGATGATGGGAGGAGTTCTATTTATACACACGGAACATCCGTAAGTGTAGCAATGAG
 -----+-----+-----+-----+-----+-----+-----+
 10620 GTACTACTACCCCTCAAGATAATATGTGTGCCCTTGTAGGCCATTTACATCGTTACTC

b M M M G G V L L Y T R K H P *

10621 AAATGATAATATCGAFACTCTGAGCGATACTTTTAAAGAAGTAGCACTCAAGAAGAGATTT
 -----+-----+-----+-----+-----+-----+-----+
 10680 TTTTACTATATAGCTATGAGACTCGCTATGAAATCTTTCATCGTGAGTCTCTCTAAA

10681 AAGTTTACTTGGTGAACAGTTTTCTTCGCCAAGTAAACCACCATTTGAAAGGGGAGATG
 -----+-----+-----+-----+-----+-----+-----+
 10740 TTCAAATGAACCACCTTTGTCAAAAAGAAGCGGTTTCATTTGGTGGTAACTTTCCCCCTCTAC

10741 TTTTTCGAAAACCTTGCACAGAAAAAAGGATTAATTTATGTCATGTGTAATTCATTACATGTC
 -----+-----+-----+-----+-----+-----+-----+
 10800 AAAAGCTTTTGAACGCTGCTTTTTCCTAATAATAACAGTACACATTAAGTAATGTAACG

10801 TCACAGTTGATTTTAAAGAGATATGAATAAGGAGAAATCATGAAATCAATCAACAATTTT
 -----+-----+-----+-----+-----+-----+-----+
 10860 AGTGTCAACTAAAATTCCTATACTTATTCCTCTTTAGTACTTTAGTTAGTTGTTTAAA

c TAACAAATGCTTGCTGCTTATTACTGACAGCGAGTAGCCCTGTTTTTCAGCTGCAACAGTTT

M K S I N K F L - orf4_670, homologue of sp0463, LPX1G

PCT/US05/27239

277/487

Figure 101Z

```

10861 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+ 10920
      ATTGTTACGAAACGCGGAATAATGACTGTGCTCGCTCATCGGACAAAAGTCGACGTTGTCAAA
c      T M L A A L L L T A S S L F S A A T V F -
      TTGGCGGGACAAATGTTAGTACAGCACCAGATGCTGTACTAAAACCTTTAAACAATCCATA 10980
10921 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      AACGCCCGCTGTTACAATCATGTCGTCGCTACGACAATGATTTTGAATTTGTTAGGTAT
c      A A D N V S T A P D A V T K T L T I H K -
      AGTTACTGCTCAGAAAGATGATTTAAAGACTTGGGATACAAACGGTCCTAAAGGATATG 11040
10981 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TCAATGACGAGAGCTTCTACTAAATTTCTGAACCCCTATGTTTGCACGAGATTTCCCTATAC
c      L L L S E D D L K T W D T N G P K G Y D -
      ATGGAACCTCAATCTAGTTTAAAGATTTAACTGGAGTTGTAGCTGAGGAAATCCAAATG 11100
11041 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TACCTTGAGTTAGATCAAAATTTTCTAAATTTGACCTCAACATCGACTCCCTTTAAGGTTTAC
c      G T Q S S L K D L T G V V A E E I P N V -
      TATACCTTGAATTACAAAAGTATAATTTGACTGATGGTAAGGAAAAGAAAATCTTAAAG 11160
11101 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      ATATGAAACTTAATGTTTTCATATTAACACTGACTACCATTCCCTTTTCTTTTGAATTTTC
c      Y F E L Q K Y N L T D G K E K E N L K D -
      ATGATAGTAAATGGACAACAGTTCATGGTGGTTTGACAACATAAAGATGGACTTAAATTTG 11220
11161 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TACTATCATTTACCTGTTGTCAAGTACCACCAAACTGTGATTTCTACCTGAATTTTAAAC
c      D S K W T T V H G G L T T K D G L K I E -
      AAACCAGTACTCTTAAAGGTGTATTCGTATTCGTGAGGATAGAACAAAGACTACCTATG 11280
11221 -----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
      TTTGGTCAATGAGAATTTCCACACATAGCATAAAGCCTCTATCTTTGTTCTGATGGATAC
c      T S T L K G V Y R I R E D R T K T T Y V -

```

PCT/US05/27239

278/487

Figure 101AA

11281 TTGGTCCCTAATGGGCAAGTATTAACAGGTTCAAAAGCCGTACCTGCTCTTGTAACCTCTTC 11340
 -----+-----+-----+-----+-----+-----+-----+-----+
 AACAGGATTACCCGTTCAATAATTGTCACCAAGTTTTCGGCATGGACGAGAACATGAGAAG

C G P N G Q V L T G S K A V P A L V T L P -

11341 CACTTGTTAAACAATAATGGGTACAGTAATTGATGCACATGTTTCCCTAAAATAATCATATA 11400
 -----+-----+-----+-----+-----+-----+-----+
 GTGAACAATTGTTATTACCATGTCATTAACACTACGGTACAAAAGGGATTTTAAAGTATAT

C L V N N N G T V I D A H V F P K N S Y N -

11401 ATAAACCAGTTGTAGATAAAAGAATTGCTGATACCTTTGAATTATAACGATCAAAAATGGTC 11460
 -----+-----+-----+-----+-----+-----+-----+
 TATTGGTCAACATCTATTTCTTAACGACTATGAACCTTAATAATTTGCTAGTTTACCAG

C K P V V D K R I A D T L N Y N D Q N G L -

11461 TGTCTATCGGTACTAAAATCCCATATGTTGTTAATACAAACAATCCCAAGTAATGCAACAT 11520
 -----+-----+-----+-----+-----+-----+-----+
 ACAGATAGCCATGATTTTAGGGTATACAACAATAATGTTGTTAAGGTTCAATTACGGTTGTA

C S I G T K I P Y V V N T T I P S N A T F -

11521 TTGCAACTTTCATTTGGTCAGATGAAATGACAGAAAGGCTCTAACTTATAATGAAGATGTAA 11580
 -----+-----+-----+-----+-----+-----+-----+
 AACGTTGAAGTAAACCAGTCTACTTTACTGTCTTCCAGATTGAATATTACTTCTACATT

C A T S F W S D E M T E G L T Y N E D V T -

11581 CAATTACTTTGAAATAATGATGCTATGGATCAAGCTGATTAATGAAGTCACTAAAGGAAATA 11640
 -----+-----+-----+-----+-----+-----+-----+
 GTTAATGAAAACCTTATTACATCGATACCTAGTTGGACTAATAACTTCAGTGATTTCCCTTAT

C I T L N N V A M D Q A D Y E V T K G N N -

11641 ATGGCTTTAACTTAAAATAACAGAAAGCAGGTTTAGCTAAAATAATGTTAAGGATGCAG 11700
 -----+-----+-----+-----+-----+-----+-----+
 TACCGAAATTGAAATTTAATTGCTCTCCGTCCTCCAAAATCGATTTAATFACCATTCCCTACGTC

PCT/US05/27239 279/487

Figure 101AB

C G F N L K L T E A G L A K I N G K D A D -
 ACCAAAAATCCAAATTACTTACTCAGTACTTGAACCTCAGTCTGTTGCCAGACATTC
 11701 -----+-----+-----+-----+-----+ 11760
 TGGTTTTTTAGGTTTAAATGATGAGTCGATGAAACTTGAAGTGAACGACACAGTCTGTAAAG
 C Q K I Q I T Y S A T L N S L A V A D I P -
 CTGAAAGTAACGATATTACATATCATACGGAAATCATCAAGATCATGGGAATACTCCAA
 11761 -----+-----+-----+-----+-----+ 11820
 GACTTTCAATTCCTATAATGATAGTAATGACCTTTAGTAGTTCTAGTACCCCTTATGAGGTT
 C E S N D I T Y H Y G N H Q D H G N T P K -
 AACCAACTAAACCTAATAATGGTCAAAATTAACAGTAACATAAGACATGGGACAGTCAACCTG
 11821 -----+-----+-----+-----+-----+ 11880
 TTGGTTGATTTGGATTATTACCAGTTTAAATGTCATTGATTCCTGTACCCCTGTGTCAGTTGGAC
 C P T K P N N G Q I T V T K T W D S Q P A -
 CTCCTGAGGGGGTGAAGCGACTGTTCAACTTGTAAATGCCAAGACTGGTGAGAAAGTCCG
 11881 -----+-----+-----+-----+-----+ 11940
 GAGGACTCCCCACTTTCGCTGACAAGTTGAACATTTACGGTTCGACCACCTCTTTCAGC
 C P E G V K A T V Q L V N A K T G E K V G -
 GTGCTCCTGTAGAACCTTTCAGAAATAATGGACATATACTTGGAGTGGTCTAGATAATT
 11941 -----+-----+-----+-----+-----+ 12000
 CACGAGGACATCTTGAAAGTCTTTTATTAACCTGTATATGAACCTCACCCAGATCTATATA
 C A P V E L S E N N W T Y T W S G L D N S -
 CTATTGAATACAAAGTTGAAGAATAATAATGGATACTCAGCTGAATACACAGTAGAGA
 12001 -----+-----+-----+-----+-----+ 12060
 GATAACTTATGTTTCAACTTCTTCTTATATTAACCTATGAGTCCGACTTATGTCATCTCT
 C I E Y K V E E E Y N G Y S A E Y T V E S -
 GCAAAGGGAAGTTGGGGTAAATAAAGTGAAGATAATAACCCAGCTCCAATCAATCCTG
 12061 -----+-----+-----+-----+-----+ 12120

280/487

Figure 101AC

CGTTTCCCTCAACCCCAATTTTGGACCTTTCTATTTATTTGGTCCGAGGTTAGTTAGGAC
 K G K L G V K N W K D N N P A P I N P E -
 12121 AAGAACCCAGGTGTAATAACATACGGTAAAAAGTTTGTCAAAAGTAGACCACAAAAGATACTC
 TTTCTGGTGCACATTTTGTATGTCCTTTTCAAAACAGTTTCACTCTGGTTTCTATGAG 12180
 E P R V K T Y G K K F V K V D Q K D T R -
 12181 GTCTAGAAAAATGCCAGTTCGTTGTTAAAAAAGCAGATAGCAATAATAATATGTCCTTTA
 CAGATCTTTTACGGTCAAGCAACAATTTTTCGTCATCTGTTATTTATATAACGGAAAT 12240
 L E N A Q F V V K K A D S N K Y I A F K -
 12241 AGTCAACTGCACAACAGCTGCAGATGAAAAGCAGCAGCAACTGCAAAACAAAATTTGG
 TCAGTTGACGTTGTTGTCGACGTCATTTTTCGTCGTCGTTGACGTTTGTGTTTAAACC 12300
 S T A Q Q A A D E K A A A T A K Q K L D -
 12301 ATGCAGCGGTAGCAGCTTACACAAATGCTGCAGATAAAGCAAGCCGCTCAAGCTCTAGTAG
 TACGTGCCCATCGTCAATGTTTACGACGTCATTTTCGTTCCGGCAGTTCCGAGATCATC 12360
 A A V A A Y T N A A D K Q A A Q A L V D -
 12361 ATCAAGCACAGCAAGAATACAAATGCTTACAAAGAAGCCAAAATTTGGTTATGTTGAAG
 TAGTTCCGTTGCTTATGTTTACATCGAATGTTTCTTCGTTTAAACCAATACAACTTC
 Q A Q Q E Y N V A Y K E A K F G Y V E V -
 12421 TAGCTGAAAAGATGAAGCAATGGTCTTACTTCTAATACGGATGGTCAATTCCAAATTT
 ATCGACCTTTTCTACTTCCGTTACCAAGAATGAAGATTATGCCCTACCAGTTAAGGTTTAAA 12480
 A G K D E A M V L T S N T D G Q F Q I S -

PCT/US05/27239 82/487

Figure 101AE

```

a      V M A L C F S L V W G A H A V Q A Q E D -
      CACACGTTGGTCTTGC AATTGGAGAACTATCAGGAGGTTAGTCAATTGCCATCTCGT
12901 -----+-----+-----+-----+-----+-----+-----+-----+
      GTGTGCAACCAGAACGTTAACCTCTTGATAGTCCCTCCACCAATCAGTTAACGGTAGAGCA
      H T L V L Q L E N Y Q E V V S Q L P S R -
      GATGGTCATCGGTTGCAAGTATGGAAGTTGGATGATTCGTAATCCCTATGATGATCGGGTG
12961 -----+-----+-----+-----+-----+-----+-----+-----+
      CTACCAGTAGCCAACGTTTCATATACCTTCAACCTACTAAGCATAAGGATACTACTAGCCCCAC
      D G H R L Q V W K L D D S Y S Y D D R V -
      CAAATGTAAAGACTTGCATTCGGTGGATGAGATAAACTTCTTCTTCAAAAAGACT
13021 -----+-----+-----+-----+-----+-----+-----+-----+
      GTTTAACATTCCTGAAAGTAAAGCACCCCTACTCTTATTTGAAAGAGAAGTTTTTCTGA
      Q I V R D L H S W D E N K L S S F K K T -
      TCGTTTGAGATGACCTTCCCTTGAGAAATCAGATGAAAGTATCTCATATCCAAATGGTCTT
13081 -----+-----+-----+-----+-----+-----+-----+-----+
      AGCAAACCTCTACTGGAAGAACTCTTAGTCTAACTTCAATAGAGTATAAGGTTTACCAGAA
      S F E M T F L E N Q I E V S H I P N G L -
      TACTATGTTCCGTCATATATCCAGACGGATGCGGTTTCTTATCCAGCTGAATTTCTTTT
13141 -----+-----+-----+-----+-----+-----+-----+-----+
      ATGATACAAGCGAGATAATAGGTCGCTCCCTACGCCAAAGAAATAGGTCGACTTAAAGAAAAA
      Y Y V R S I I Q T D A V S Y P A E F L F -
      GAAATGACAGATCAAACGGTAGAGCCCTTTGGTCAATTTAGCGAAAAAACAAGATACAATG
13201 -----+-----+-----+-----+-----+-----+-----+-----+
      CTTTACTGTCAGTTTGGCCATCTCGGAAACCAGTAAACATCGCTTTTCTTCTATGTTAC
      E M T D Q T V E P L V I V A K K T D T M -
      ACAACAAAGGTGAAGCTGATAAAGGTGGATCAAGACCACAATCGCTTGGAGGGTGTGGC
13261 -----+-----+-----+-----+-----+-----+-----+-----+

```


Figure 101AF

```

a      T T K V K L I K V D Q D H N R L E G V G -
      TGTGTTCCACTTCGACTAATTCACCTAGTTCGGTGTAGCGAACCTCCACAGCCG
      TTTAAATGGTATCAGTAGCAAGAGATGGTTCGAAAAAGAGGTTCCCTTGATGGAGAA 13380
      AAATTTAACCATAGTCATCGTTCTCTACCAAGACTTTTCTCCAAGGGAACATAACCTCTT
a      F K L V S V A R D G S E K E V P L I G E -
      TACCGTTACAGTTCCTTCTGGTCAAGTAGGGAGAACTCTATACATGATAAAAATGGAGAG
      ATGGCAATGTC AAGAAGACCAAGTTCATCCCTCTTGAGAGATATGACTATTTTACCTCTC 13440
a      Y R Y S S S G Q V G R T L Y T D K N G E -
      ATTTTGTGACAAAATCTCTCTTGGGAACATACTGTTTCAAGGAGTGGAGCCACTGGCA
      TAAAAACACTGTTTAGAAGGAGAACCTTGTATAGCAAGTTCCTCCACCTCGGTGACCGT 13500
a      I F V T N L P L G N Y R F K E V E P L A -
      GGCTATGCTGTACGACGCTGGATACGGATGTCCAGCTGGTAGATCATCAGCTGGTGACG
      CCGATACGACAAATGCTGGACCTATGCCCTACAGGTCGACCATCTAGTAGTCGACCCTGC 13560
a      G Y A V T T L D T D V Q L V D H Q L V T -
      ATTACGGTTGTCAATCAGAAAATTACCACGTTGGCAATGTTGACTTTTATGAAGGTGGATGGT
      TAATGCCAACACAGTTAGTCTTTAATGGTGCACCCGTTACAACCTGAAATACTCCACCTACCA 13620
a      I T V V N Q K L P R G N V D F M K V D G -
      CCGACCAATACCTCTCTCAAGGGCAATGTTCAAGTCATGAAGAAGAAGCGGACAC
      GCCTGGTTATGGAGAGAAGTTCCTCCGTTACAAGTTTCAGTACTTTCTTCTTTCGCCCTGTG 13680
a      R T N T S L Q G A M F K V M K E E S G H -

```

Figure 101AG

```

13681 TATACTCCTGTTCTTCAAATGGTAAGGAGTAGTTGTAACATCAGGGAAGAATGGTCTGT
-----+-----+-----+-----+-----+-----+-----+-----+
a ATATGAGGACAAGAAGTTTACCATTCCCTTCATCAACATTTGTAGTCCCTTCTTACCAGCA
13740 Y T P V L Q N G K E V V T S G K D G R -
-----+-----+-----+-----+-----+-----+-----+-----+
13741 TTCCGAGTGGAAAGGCTAGAGTATGGACATACATATTTATGGGAGCTCCAAGCTCCAAC
-----+-----+-----+-----+-----+-----+-----+-----+
a AAGGCTCACCTTCCAGATCTCATACCCCTGTATGATAAATAACCCCTCGAGGTTCCGAGTTGA
13800 F R V E G L E Y G T Y L W E L Q A P T -
-----+-----+-----+-----+-----+-----+-----+-----+
13801 GGTATGTTCAATTAACATCGCCCTGTTCCCTTACAATCGGGAAGAATACTCGTAAGGAA
-----+-----+-----+-----+-----+-----+-----+-----+
a CCAATACAAGTTAATTGTAGCGGACAAGGAAATGTTAGCCCTTCTATGAGCATTCCCTT
13860 G Y V Q L T S P V S F T I G K D T R K E -
-----+-----+-----+-----+-----+-----+-----+-----+
13861 CTGGTAACAGTGGTTAAAATAACAAGCCAGGATTTGATGTGCCAGATACAGGGGAA
-----+-----+-----+-----+-----+-----+-----+-----+
a GACCATTGTCACCAATTTTATTTGTTCCCTGCTGCTGCTTACCTACACGGTCTATGTCCTT
13920 L V T V V K N N K R P R I D V P D T G E -
-----+-----+-----+-----+-----+-----+-----+-----+
13921 GAAACCTTGATATCTTGATGCTTGTGCCATTTGTTGTTGTTGTTGTTGTTATTTATCTT
-----+-----+-----+-----+-----+-----+-----+-----+
a CTTTGGAACATATAGAACTACGAACAACGGTAAAACAACAACCAATCACCACCAATAATAGAA
13980 E T L Y I L M L V A I L L F G S G Y Y L -
-----+-----+-----+-----+-----+-----+-----+-----+
13981 ACGAAAAAACCAATTAACGTGATATTCATGTACATCATATGAAAAAGATAGCAGGCTGA
-----+-----+-----+-----+-----+-----+-----+-----+
a TCGTTTTTTGGTTTATGACTAFAAGTTACATGTAATAACTTTTCTATCGTCCGACT
14040 T K K P N N *
-----+-----+-----+-----+-----+-----+-----+-----+
14041 AGGGAAGACCAGACTCTGAGGTGATGTTAATCAGGAATCATGGTATGGCATGAA
-----+-----+-----+-----+-----+-----+-----+-----+
a TCCCTTCTGGTCTCATGAGACTCCACTACAATTAGTCCCTTAGTACCACCTACACCCGACTT
14100

```

Figure 101AH

14101 TCACAAATAACGGATATGAGGCTGGCCAGATTGTGCCAGCCCTCATTTGTGGTTATTGTTTG + 14160
 AGTGTATTATTCCTTACTCCGACCCCGTCTAACACCGGTCGGAGTAACACCCCAATAACAAC

14161 TAAACGATAGGACTGGTCTGGTAATCATTTTAGGAATGGACAGGACTGGGATCTGATT + 14220
 ATTTTGCTATCCTGACCCAGACCAATTAGTAAATAATCCTTACCTGTCTGACCCCTAAGACTAA

14221 TAAAAATGGATGGTGAATCAGAAAGAATAAGAGATTTTCTCGTTTCTCTTAGCAGATAGGAT + 14280
 ATTTTACCCTACCACITTAGTCTTTCTTTTACTCTAAAGAGCAAAAGAGAATCGTCTATCCCTA

14281 TGTCTGTTAGGAAAAGCGATAAAATGATGAGTTTGAAGATAAAGGATGCTGATAAAAAT + 14340
 ACAGACAATCCTTTTCCGCTATTTTACTACTCAAACCTTCTATTCCCTACGACTATTTTA

14341 GGTAACAAACAAAAGCAAAAACGAAATAATCTCCTATTAGGAGTGGTATTTTTCATTTGG + 14400
 CCATTTTGTGTTTTTTCGTTTTTTCGTTTATTAGAGGATAAATCCTCACCATAAAAAGTAACC

b V K T K K Q K R N N L L L G V V F F I G -

14401 AATGGCGGTAATGGCGTATCCGCTGGTGTCTCGCTTGTATTATCGAGTGGAAATCAATCA + 14460
 TTACCGCCATFACCGCATAGGGCACCACAGAGGGAACATAATAGCTCACCTTAGTTTAGT

b M A V M A Y P L V S R L Y Y R V E S N Q -

14461 ACAAAATGCTGACTTTGATAAGGAAAAGCAACGGTTGGATGAGGCTGACATGATGAACG + 14520
 TGTTTAACGACTGAAACTATTCCTTTTTCGTTGCAACCTACTCCGACTGTAACCTACTTGC

b Q I A D F D K E K A T L D E A D I D E R -

AATGAAATTGGCACAAAGCCTTCAATGACTCTTTGAATAATGTAGTGAGTGGCGATCCTTG

M L I K M -orf6_670; homologue of sp0466, sortase

PCT/US05/27239

286/487

Figure 101AI

14521 -----+-----+-----+-----+-----+-----+-----+-----+ 14580
TTACTTTAACCGTGTTCGGAACTTACTGAGAACTTATACATCACTCACCCTAGGAAC
b M K L A Q A F N D S L N N V V S G D P W -
14581 -----+-----+-----+-----+-----+-----+-----+ 14640
GTCGGAAGAAATGAAGAAAAAGGGCCGAGAGTATGCACGTATGTTAGAAAATCCATGA
CAGCCTTCTTTTACTTCTTTTCCCGCTCCTCATACGTGCATACAATCTTTAGGTACT
b S E E M K K K R A E Y A R M L E I H E -
14641 -----+-----+-----+-----+-----+-----+-----+ 14700
GGGGATGGGGCATGTGGAAAATCCCGTTATTGACGGTGGATTTGCCGGTTTATGCTGGTAC
CGCCTACCCCGTACACCTTTAGGGCCATAACTGCACCTAAACGGCCAAAATACGACCATG
b R M G H V E I P V I D V D L P V Y A G T -
14701 -----+-----+-----+-----+-----+-----+-----+ 14760
TGCTGAAGAGGATTTGCAGCAAGGGGCTGGGCATCTAGAGGGAACCTCTCTGCCGATCGG
ACGACTTCTCCATAACGTCTTCCCGACCCCGTAGATCTCCCTTGAAGAGACGGCTAGCC
b A E E V L Q Q G A G H L E G T S L P I G -
14761 -----+-----+-----+-----+-----+-----+-----+ 14820
AGGCAATTCCGACCCATGCGGTGATTACGGCACATACAGGTTGCCAACAGCTAAGATGTT
TCCGTTAAGCTGGGTACGCCACATAATGCCGTGATGTCCTCCAAACGGTTGTCGATCTACAA
b G N S T H A V I T A H T G L P T A K M F -
14821 -----+-----+-----+-----+-----+-----+-----+ 14880
TACGGATTTGACCAAACTTAAAGTTGGGATAAGTTTATGTCACAAATATCAAGGAAGT
ATGCCATAAAGTGGTTTGAATTTCAACCCCTATTCAAAAATACACGTTGTTATAGTTCCTTCA
b T D L T K L K V G D K F Y V H N I K E V -
14881 -----+-----+-----+-----+-----+-----+-----+ 14940
GATGCCATCAAGTGGATCAAGTAAAGGTGATTGAGCCGCAACTTTGATGATTTATT
CTACCGGATAGTTCACCTAGTTCATTTCCACTAATCCGCTGCTTGAACACTACTAATAAA
b M A Y Q V D Q V K V I E P T N F D D L L -

PCT/US05/27239

287/487

Figure 101A

14941 GATTGTACCAGGTCATGATTAATGACCTTGGCTGACTTGTACGCCATACATGATCAATAC + 15000
 CTAACATGGTCCAGTACTAATACTACTGGAACGACTGAACATGCGGTATGTACTAGTTATG
 b I V P G H D Y V T L L T C T P Y M I N T -

15001 CCATCGTCTATTGGTTCGGGGCATCGGATACCGTACGTAGCAGAGGTTGAGGAAGATT + 15060
 GGTAGCAGATAACCAAGCCCCGTAGCCCTATGGCATGCATCGTCTCCAACCTCTCTTAA
 b H R L L V R G H R I P Y V A E V E E F -

15061 TATTGCAGCAAAACACTCAGTCTCTATCGCTACCTGTTTATGTGGCAGTTGGTTT + 15120
 ATAACGTCGTTTGTGAGTCAGTAGAGATAGGATGGACAAATAATACACCCGTCACCCAAA
 b I A A N K L S H L Y R Y L F Y V A V G L -

15121 GATTGTGATCTTTTATGGATTATTCGACGGCTTGGCAAGAAAGAAAACACCGGAAA + 15180
 CTAACACTAAGAAAATACCTAATAAGCTGGCAACGGCTCTCTTTTGTGGCCTTTT
 b I V I L L W I I R R L R K K K Q P E K -

15181 GGCTTTGAAGGCGCTGAAAGCAGCAAGGAAGTGAAGGTGGAGGATGGACACAGTA + 15240
 CCGAAACTCCGGGACTTTCGTCGTTCCCTTCCCTTCCACTTCCACCTCCCTACCTGTGTCAT
 b A L K A L K A A R K E V K V E D G Q Q * -

15241 GACGTTCAAGAAAAGGCACAAAAGAAAGAAACATCCGCTGATCCCTTCTCTGATTT + 15300
 CTGCAAGTGCCTTTTTCGGTGTTTTTCCTTCTTCTTTGTAGGCGACTAGGAAGACTAAA
 b TCTTAGTAGGATTCGGCTTGGATATATCCATTTGGTGTCTCGTTATTTATTCGTATG + 15360
 AGAATCATCTAAGCGCAACCGCTATATAGGTAACACAGAGCAATAATAATAGCATAAC
 c V S R Y Y Y R I E -orf7_670, homologue of sp0467, sortase

PCT/US05/27239

288/487

Figure 101AK

15361 AGTCAAACGAGGTTATTAAGAGTTTGATGAGACGGTTCCAGATGGATAAGCAGAAC + 15420
TCAGTTGGCTCCAATAATTCTCAAACTACTCTGCCAAAGGGTCTACCTATTCCGGTCTTG
S N E V I K E F D E T V S Q M D K A E L -

15421 TTGAGGAGCGTTGGCGCTTGGCTCAAGCCCTTCAATGGACCTTGAACCACTGAAATTC + 15480
AACTCCCTCGCAACCCGCAACCGAGTTCGGAAGTTACGCTGGAACCTTTGGTAGACTTTAAG
E E R W R L A Q A F N A T L K P S E I L -

15481 TTGATCCCTTTACAGAGCAAGAGAAAAGAAAGGGCTCTCAGAAATATGCCAATAATGCTAA + 15540
AACTAGGAAAATGCTCGTCTCTTTCTTTTCTTCCGACAGTCTTTATACGGTTATACGATT
D P F T E Q E K K K G V S E Y A N M L K -

15541 AGGTCCATGAGCGGATTTGGCTATGTGAAATTCCTCGGATGTGATCAGGAAATCCGATGT + 15600
TCCAGGTACTCGCCTAACCGATACACCTTTAAGGACGCTAAGTCTCTTTAAGGCTACA
V H E R I G Y V E I P A I D Q E I P M Y -

15601 ATGTCGGAACGAGTGAGGAAATCTTTCAGAAAGGGCCGAGGATGCTAGAGGGAGCTTCGT + 15660
TACAGCCTTGCTCACTCCTTTAAGAAGTCTTCCCGGCTCCTAACGATCTCCCTCGAAGCA
V G T S E E I L Q K G A G L L E G A S L -

15661 TACCGGTTGGTGGTGAATAACCCACACAGTTGTCACTGCTCATAGAGGATTACCCGACGG + 15720
ATGGCCCAACCCACTTTTATGGTGTGTCACACAGTACAGGATCTCCCTAATGGCTGCC
P V G G E N T H T V V T A H R G L P T A -

15721 CAGAACTGTTTAGTCAATTTGGATAAGATGAAAAGGGGATGCTTTTATCTTCACGTTT + 15780
GTCTTGACAATAATCAGTTAACCTATTCTACTTTTTCCTTCCCTACAGAAAATAGAGTGCAAA

PCT/US05/27239

289/487

Figure 101A1

c E L F S Q L D K M K K G D V F Y L H V L -
 TAGACCAGGTGTTGGCCCTACCAAGTGGATCAGATTTTGACGGTTGAGCCAAATGACTTTG + 15840
 ATCTGGTCCACAACCGGATGGTTACCTAGTCTAAAACCTGCAACTGCCAAGTTCGGTTTACTGAAAC
 D Q V L A Y Q V D Q I L T V E P N D F E -
 AGCCTGCTTGTGATTCAACATGGGGAAGATTATGCGACCTTGTGACCTGTACACCGTATA + 15900
 TCGGACAGAACTAAGTTGTACCCCTTCTAATACGCTGGAACAACCTGGACATGTGGCATAT
 P V L I Q H G E D Y A T L L T C T P Y M -
 TGATTAACAGTCAATCGTCTGTTGGTACGTTGGGAAGCGGATTCGGTATACGGCACCAATTG + 15960
 ACTAATGTCTAGTAGCAGACAAACCAATGCACCCCTTCCCTAAGGCATATGCCGGTGGTTAAC
 I N S H R L L V R G K R I P Y T A P I A -
 CAGAGCGAAATCGAGCGGTGAGAGAGCGTGGCAATCTGGTTGTGGTTATTGCTAGCGG + 16020
 GTCTCGCTTAGCTCGCCACTCTCTCGCACCCCGTTAAGACCAACACCAATTAACGATCGCC
 E R N R A V R E R G Q F W L L L A A -
 CGTTGGTTATGATTCTGGTATTGAGTTACGGGGTGTATCGTCAATCGTCCGATTTGTCAAAG + 16080
 GCAACCAATACATAAGACCAATACTCAATGCCCCACATAGCAGTAGCAGCGGTAACAGTTTC
 L V M I L V L S Y G V Y R H R R I V K G -
 GGCTAGAAACAATTGGAGGAGCATGTCAAAGGCTAAGCTACAGAAATTAAGTAGGG + 16140
 CCGATCTTTTGTAAACCTCGTCTGAGTACAGTTTCCGATTCGATGCTTTAATGATCCC
 a M S K A K L Q K L L G -orf8_670, homologue of sp0468, sortase
 c L E K Q L E E H H V K G *

PCT/US05/27239

290/487

Figure 101AM

```

16141 TATTTGCTGATGCTGGTAGCATGGTGGATTCCTGTTTATTGTTTGGGGCAGATGGTGTTA
-----+-----+-----+-----+-----+-----+-----+
16200 ATAAACGACTACGACCATCGTAACCACTAAGGACAAAATAACAAAACCCGGTCTACCACAAT
-----+-----+-----+-----+-----+-----+-----+
a Y L L M L V A L V I P V Y C F G Q M V L -
-----+-----+-----+-----+-----+-----+
16201 CAGTCTTTAGGACAAGTAAAGGTCATGAGATATTTTCAGAAATCTGTGACGGCCGACAGT
-----+-----+-----+-----+-----+-----+-----+
16260 GTCAGAAATCCGTTCATTTCCAGTACTCTATAAAAGTCTTAGACACTGCCGGGTGTCA
-----+-----+-----+-----+-----+-----+-----+
a Q S L G Q V K G H E I F S E S V T A D S -
-----+-----+-----+-----+-----+-----+
16261 TACCAAGAGCAATTCGAACGGTCCGTTGATTAACAACCGCTGGATTCCGAAAATCGT
-----+-----+-----+-----+-----+-----+-----+
16320 ATGGTCTCGTTAACGTTGCCAGGAACTAATGTAGTTGCCGAACCTAAGCGTTTITAGCA
-----+-----+-----+-----+-----+-----+-----+
a Y Q E Q L Q R S L D Y N Q R L D S Q N R -
-----+-----+-----+-----+-----+-----+
16321 ATTGTAGATCCCTTTTGGCGGAAGGTATGAGGTAATACCAAGTGTCTGACGATCCT
-----+-----+-----+-----+-----+-----+-----+
16380 TAAACATCTAGGAAAACCGCCTTCCCATACTCCATTTAATGGTTCACAGACTGCTAGGA
-----+-----+-----+-----+-----+-----+-----+
a I V D P F L A E G Y E V N Y Q V S D D P -
-----+-----+-----+-----+-----+-----+
16381 GATGCAGTCTACGGCTATTGTGCGATTCCGAGTTTGGAAATCATGGAGCCAGTTTATCTA
-----+-----+-----+-----+-----+-----+-----+
16440 CTACGTCAGATCCGGATAAACAGCTAAGGCTCAACCTTTAGTACCTCGGTCAATAGAT
-----+-----+-----+-----+-----+-----+-----+
a D A V Y G Y L S I P S L E I M E P V Y L -
-----+-----+-----+-----+-----+-----+
16441 GGAGCGGATTACCATTTAGCAATGGGGTTGGCCCATGTGGATGGGACGCCCTTCTCCT
-----+-----+-----+-----+-----+-----+-----+
16500 CCTCGCCTAATGGTAGTAAATCGTTACCCCAACCGGGTACACCTACCCCTGGGAGAAGGA
-----+-----+-----+-----+-----+-----+-----+
a G A D Y H H L A M G L A H V D G T P L P -
-----+-----+-----+-----+-----+-----+
16501 GTTGAGGAAAGGGATTCTCAGTGTGCTGGCCAGCCGTCAGAACCAAGCCATGTC
-----+-----+-----+-----+-----+-----+-----+
16560 CAACTCCCTTTTCCCTAAGCAAGTCACTAACGACCCGTCGACCGTCTTGGTTCGGTACAG
-----+-----+-----+-----+-----+-----+-----+

```


Figure 101AM

```

a V E G K G I R S V I A G H R A E P S H V -
    TTTTCCGCCATTGGATCAGCTAAAGTTGGAGATGCTCTTTATTATGATAATGGCCAG
    16561 -----+-----+-----+-----+-----+ 16620
    AAAAAGCGGTAACCTAGTCGATTTTCAACCTCAGAGAAATAATACTATTACCGGTC
    F F R H L D Q L K V G D A L Y Y D N G Q -
    GAAATGTAGAAATCAGATGATGGACACAGAGATTATTTTACCCTCGGAATGGGAAAA
    16621 -----+-----+-----+-----+-----+ 16680
    CTTTAAACATCTTATAGTCTACTACCTGCTCTCTAATAAAATGGCAGCCTTACCCTTTT
    E I V E Y Q M M D T E I I L P S E W E K -
    TTAGAAATCGGTTAGCTCTAAAATAATCATGACCTTGATTAACCTCGGATCCGATCCCTACC
    16681 -----+-----+-----+-----+-----+ 16740
    AATCTTAGCCAAATCGAGATTTTATAGTACTGGAACCTATTGGACGCTAGGCTAAGGATGG
    L E S V S S K N I M T L I T C D P I P T -
    TTTAATAAACGCTTATTAGTGAATTTTGAACGAGTCGCTGTTTATCAAAAATCAGATCCA
    16741 -----+-----+-----+-----+-----+ 16800
    AAATATTGGGAATAATCACTTAAACTTGTCTCAGCGACAATAAGTTTTTAGTCTAGGT
    F N K R L L V N F E R V A V Y Q K S D P -
    CAAACAGCTGCAGTTCCGAGGTTGCTTTTACGAAAGAGACAATCTGTATCGCGTGT
    16801 -----+-----+-----+-----+-----+ 16860
    GTTTGTGACGCTCAACGCTCCCAACGAAATGCTTTCTTCCCTGTAGACATAGCGCACAA
    Q T A A V A R V A F T K E G Q S V S R V -
    GCAACCTCTCAATGGTTGTACCGTGGCTAGTGGTACTGGCATTTCTGGGAATCCCTGTTT
    16861 -----+-----+-----+-----+-----+ 16920
    CGTTGGAGAGTTACCAACATGGCCCGGATCACCATGACCGTAAAGACCCCTTAGGACAAA
    A T S Q W L Y R G L V V L A F L G I L F -
    GTTTGTGGAAGCTAGCACGTTTACTACGAGGAAATAAAAGAAATGAAAGGAAAGCTA
    16921 -----+-----+-----+-----+-----+ 16980

```

PCT/US05/27239 292/487

Figure 101AN

CAAACACCTTCGATCGCAAAATGATGCTCCCTTATTTTTCCTTACCTTTCCTTTCGAT

V L W K L A R L L R G K *

AGGCTGTTCCTTTTCCGGCTCTTTGTCACACTGTAGTGGGTGAAAAAAGCTAAGCTCG
TCCGACAAAGGAAAAAGGCCGAGAAACAGTTGACATCACCCCACTTTTTCGATTCGAGC

16981 17040

AGAAAGGACAAATTTGTCCTTCTTTTTTGATATTCAGAGCGATAAAAATCCGTTTTTT
TCTTTCCTGTTTAAACAGGAAAGAAAAAACTATAAGTCTCGCTATTTTAGGCAAAAAA

17041 17100

GAAAGTTTTTCAA

17101 -----+-- 17112

CTTCAAAGTTTT

293/487

M1, strain 2580

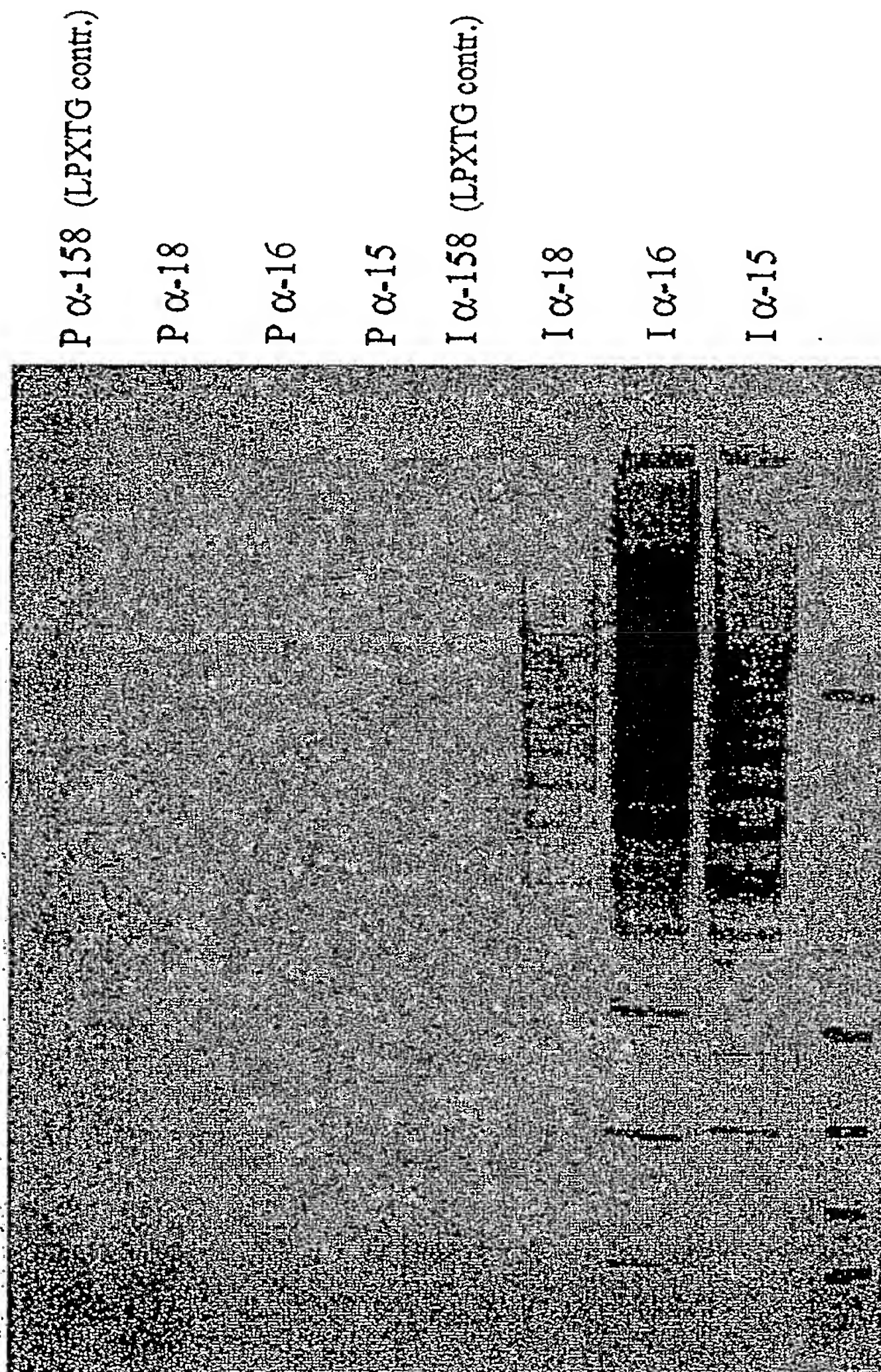


Figure 102

LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

PCT/US05/27239 294/487

M1, strain 2913

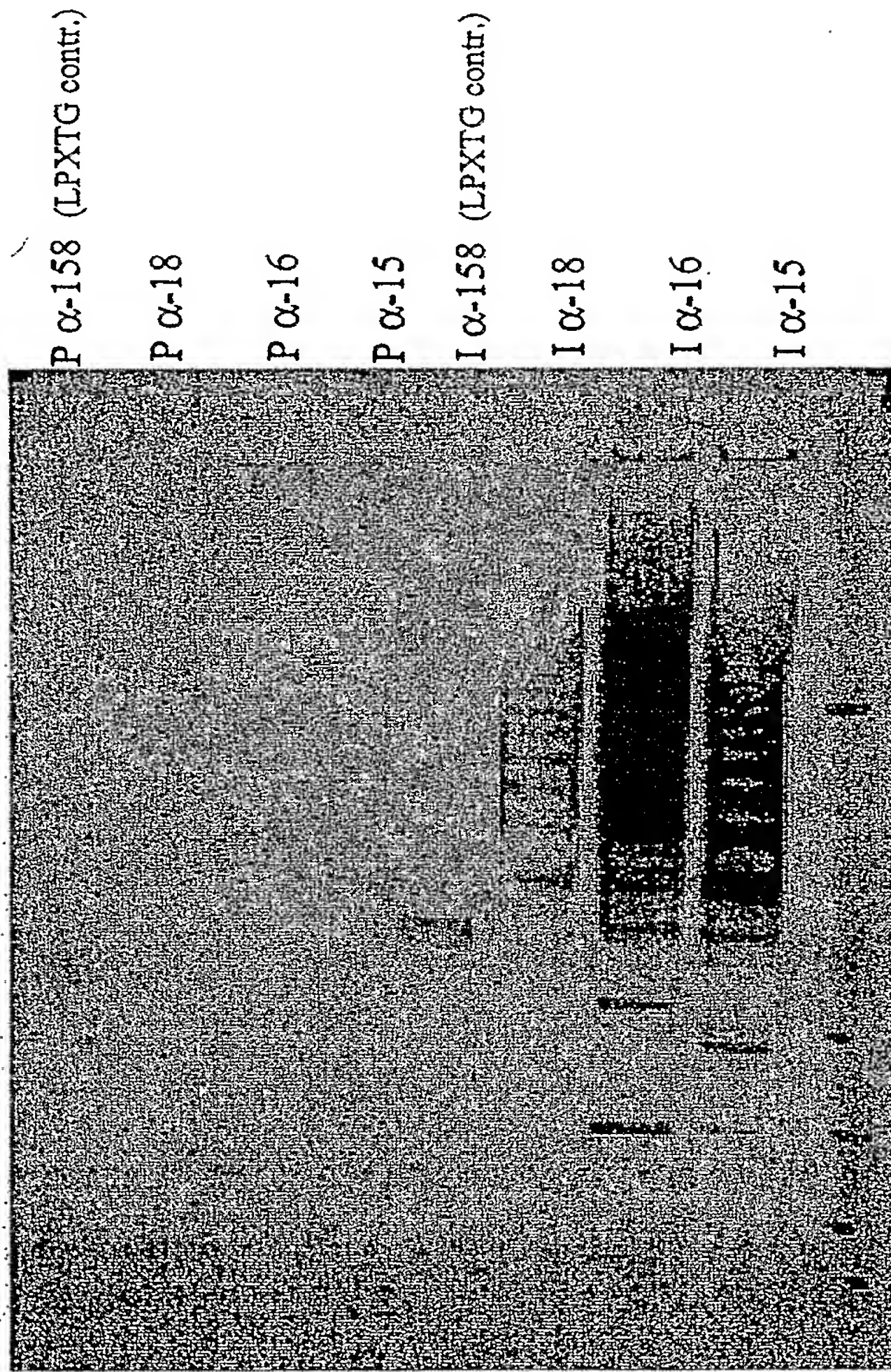


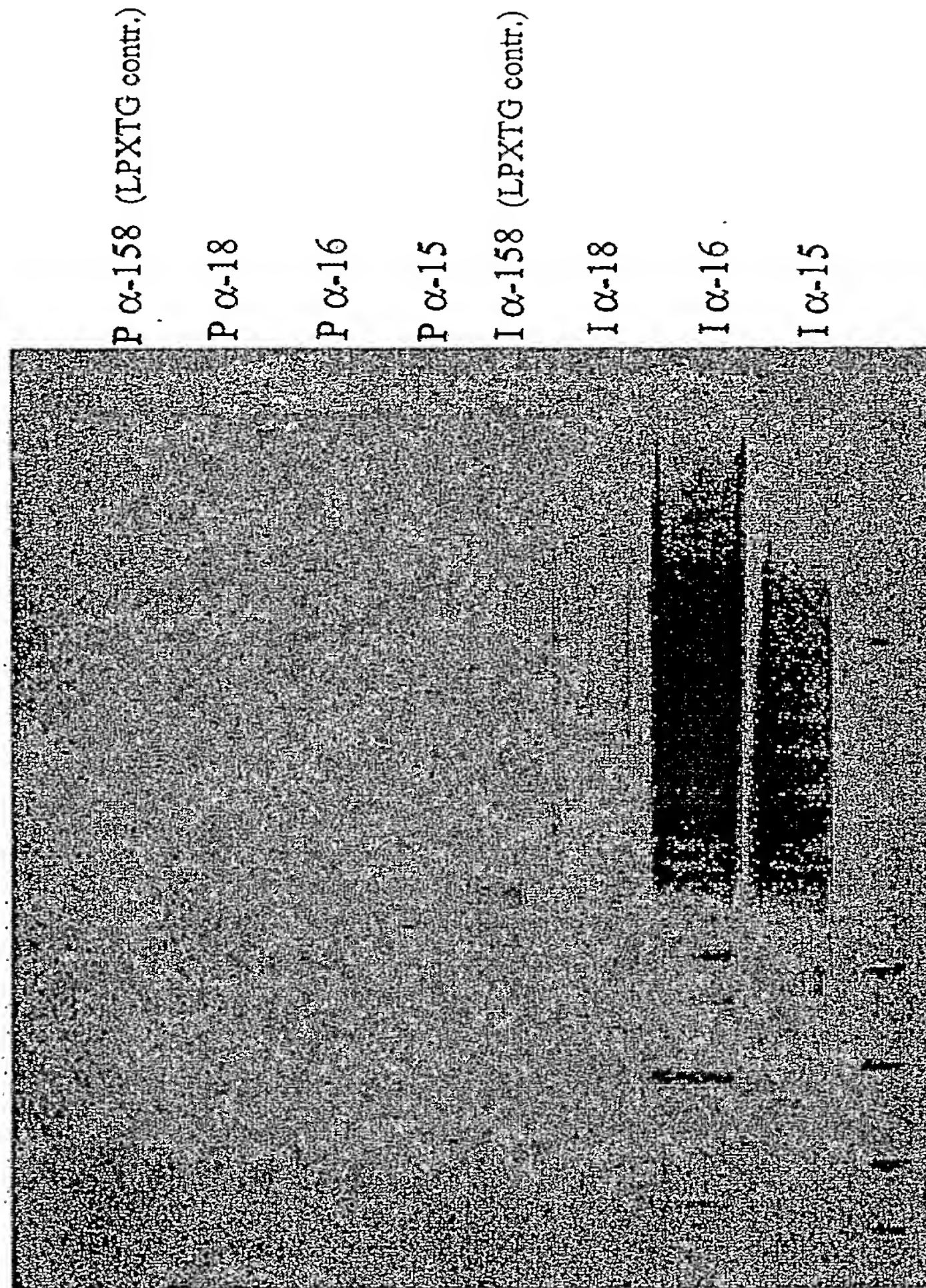
Figure 103

LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

M1, strain 3280



LEGEND:

I α-#: immune serum anti-#

P α-#: pre-immune serum anti-#

Figure 104

M1 strain 3348

P α -158 (LPXTG contr.)
P α -18
P α -16
P α -15
I α -158 (LPXTG contr.)
I α -18
I α -16
I α -15

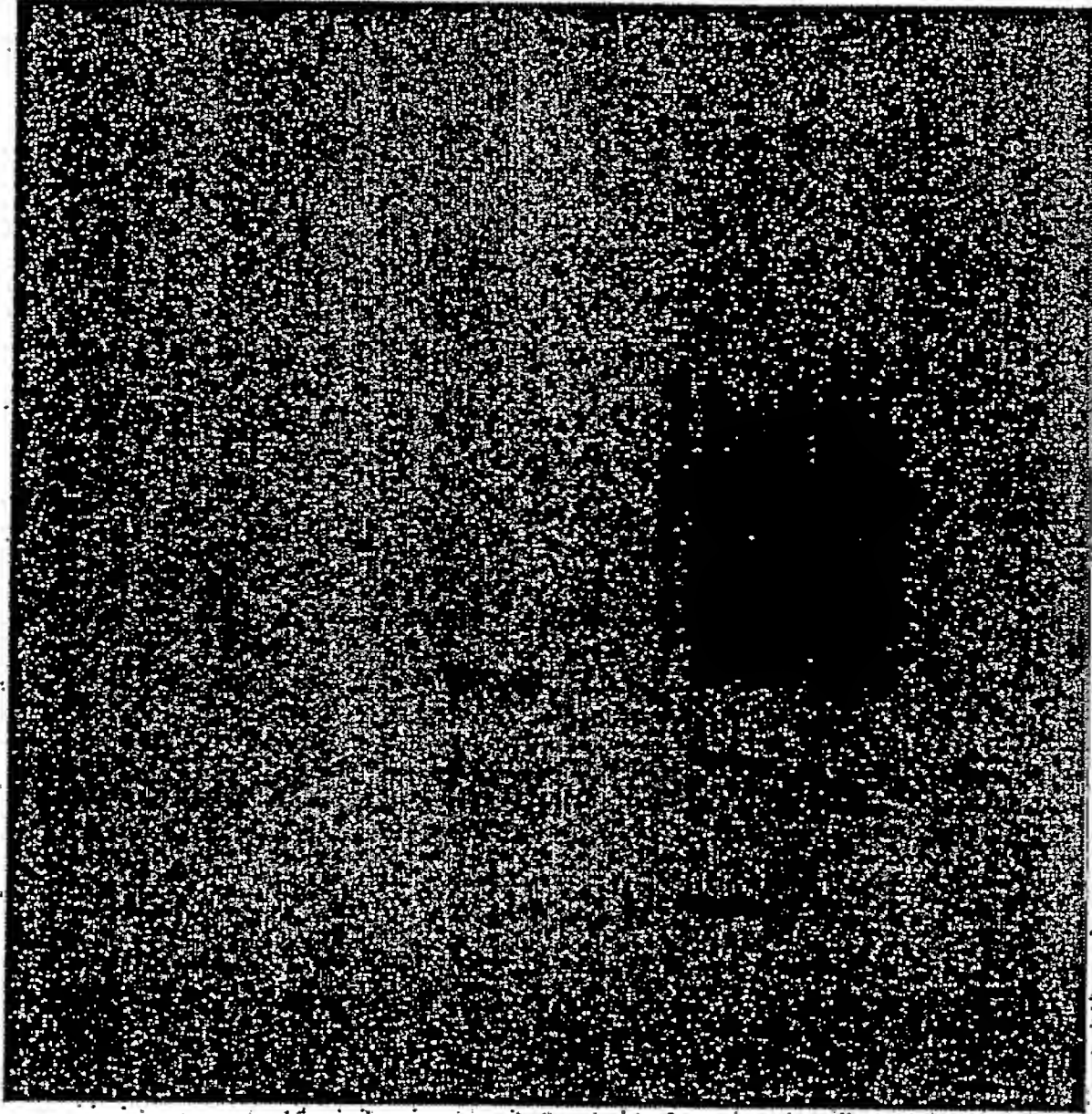


Figure 105

M1 strain 2719

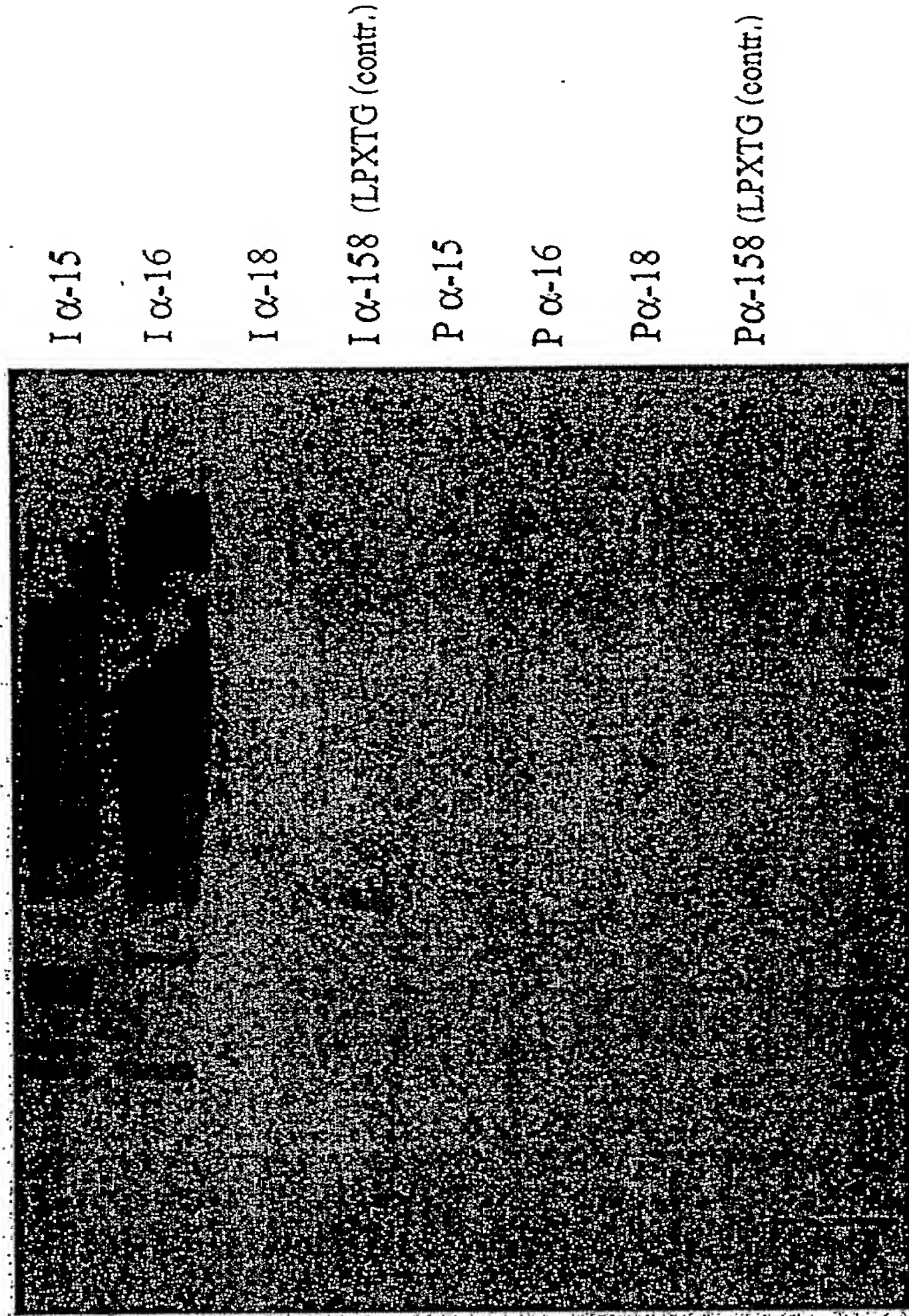


Figure 106

Western blot Western blot on fraction enriched in surface proteins of M1 (SF370)

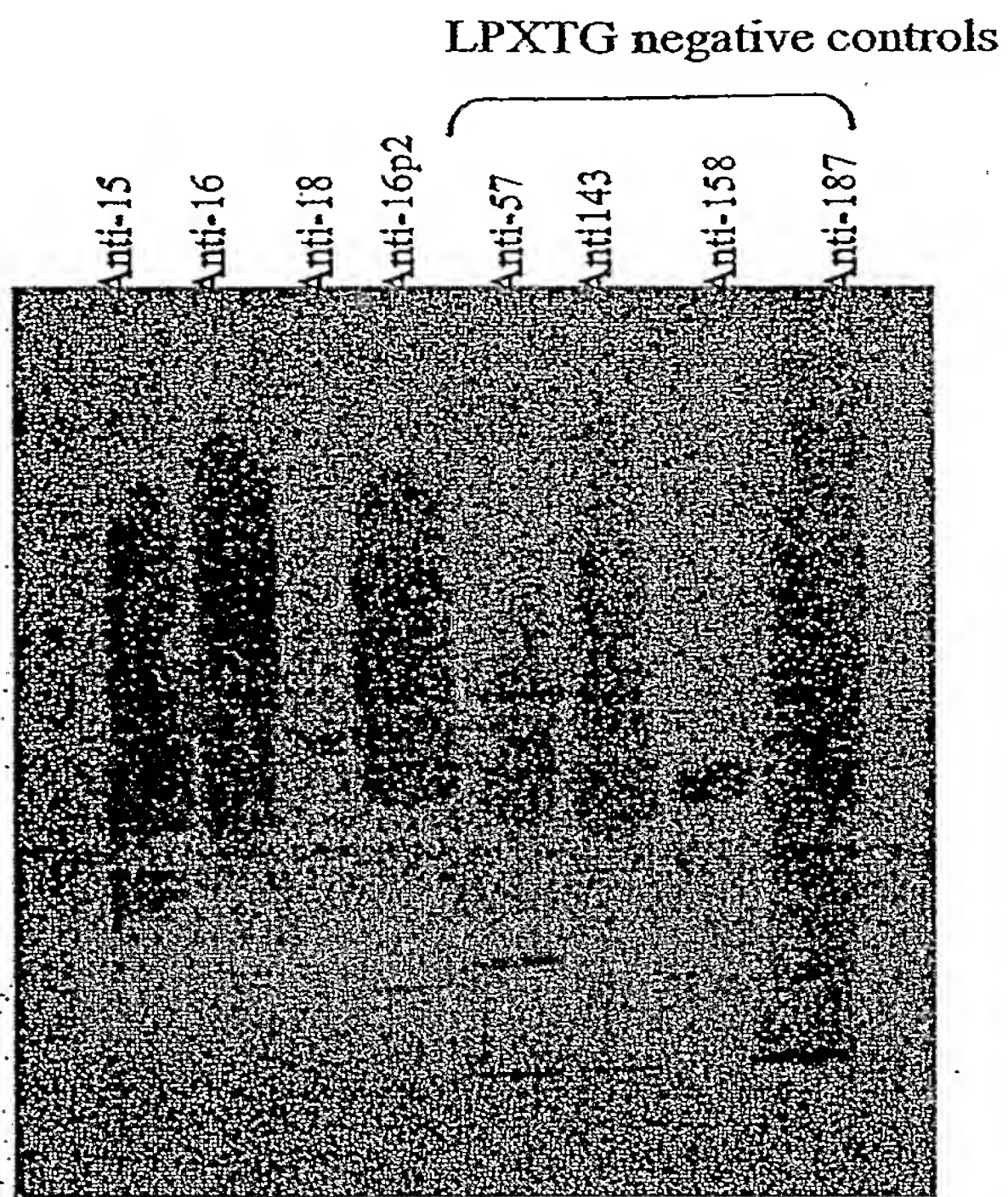
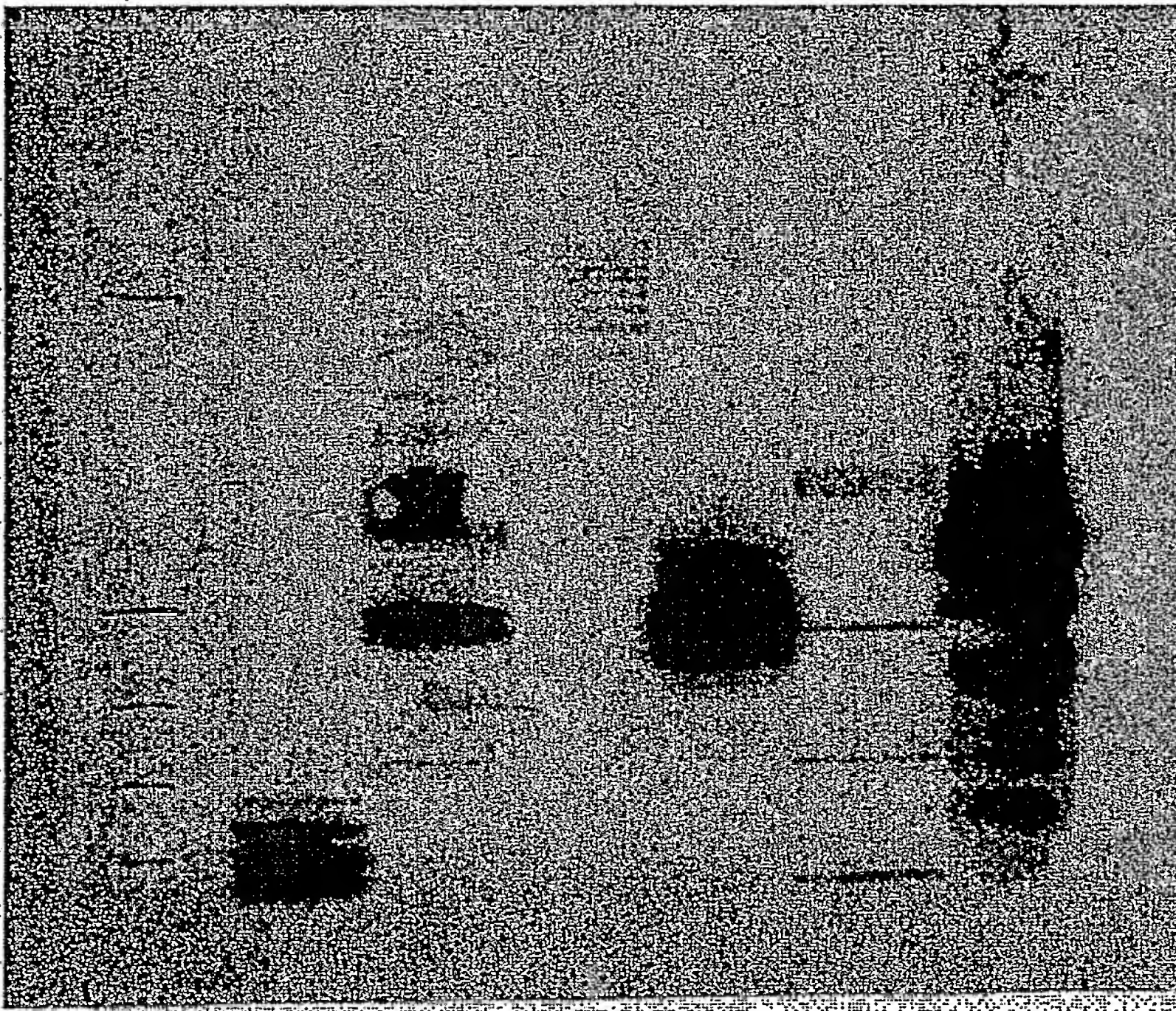


Figure 107

Western blot on fraction enriched in surface proteins of M12 (2728)

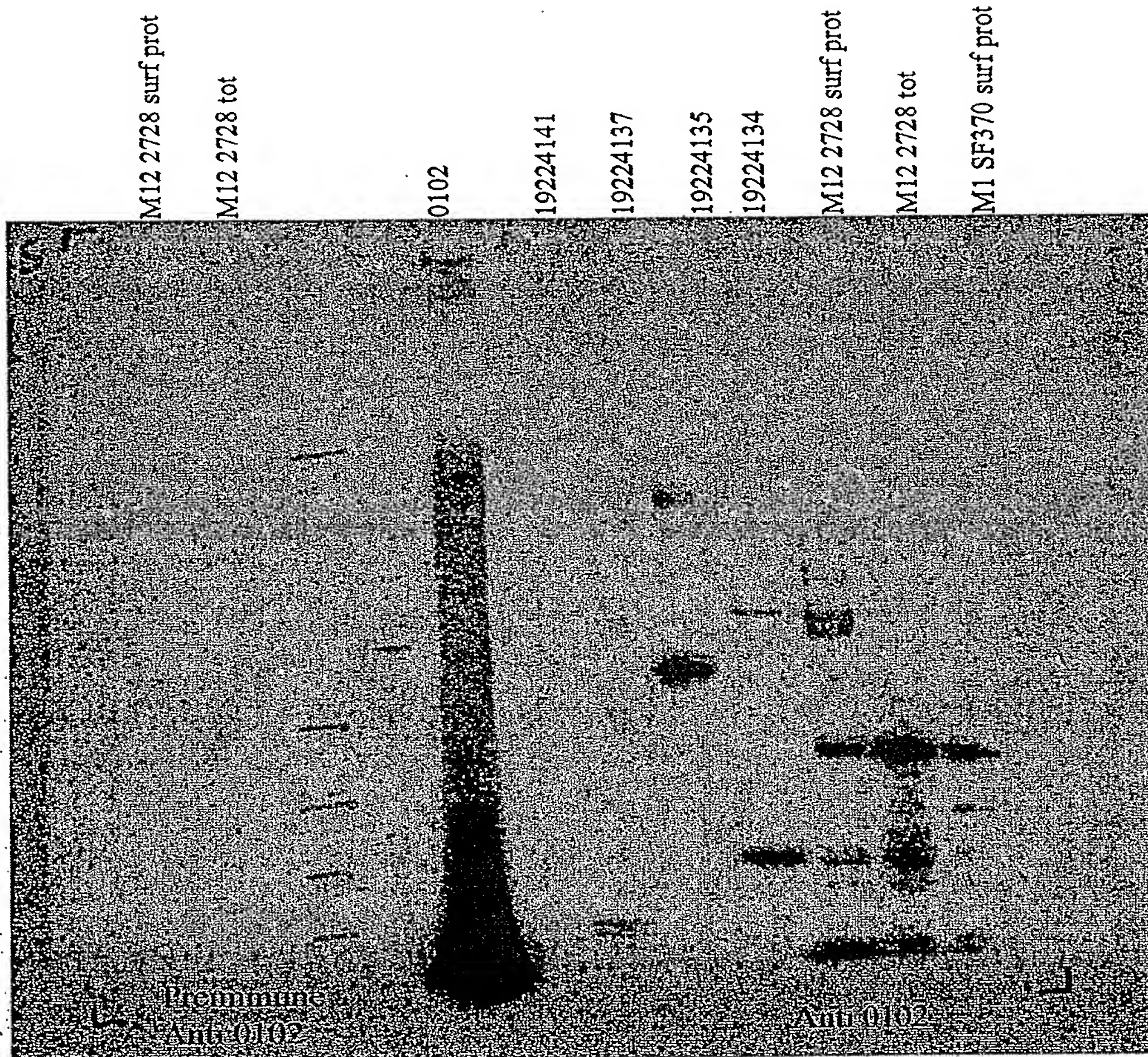
Anti-19224134
Anti-19224135
Anti-19224137
Anti-19224141
Anti-0102
Anti-158 (control)



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2a)"-, the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in M12 (strain isolate 2728). This means that in M12 three out of the five LPKTG proteins form high-molecular weight structures:

Figure 108

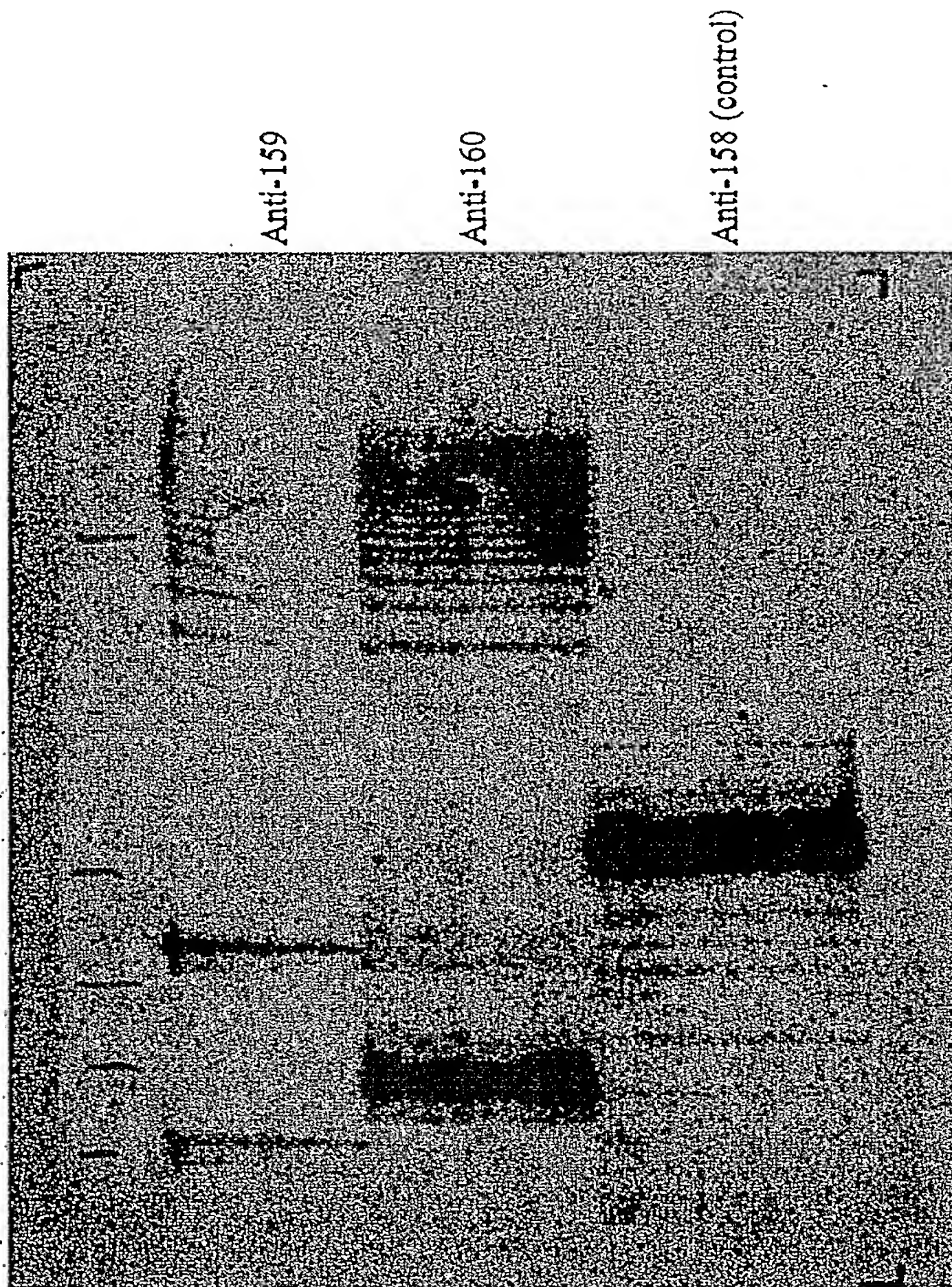
PCT/US05/27239 300/487



Note: 0102 corresponds to SpyM3_0102 in MGAS315 (M3). This protein is pretty much identical to the M12 19224139 -see also gel "WB M12 (23-11-4b2bis)", the M18 spyM18_0130 and the M5 orf 82. Here the antiserum raised against 0102 recognizes high-weight molecular structures in protein extracts from M12 (strain isolate 2728.)

Figure 109

Western blot on fraction enriched in surface proteins of M6 (2724)

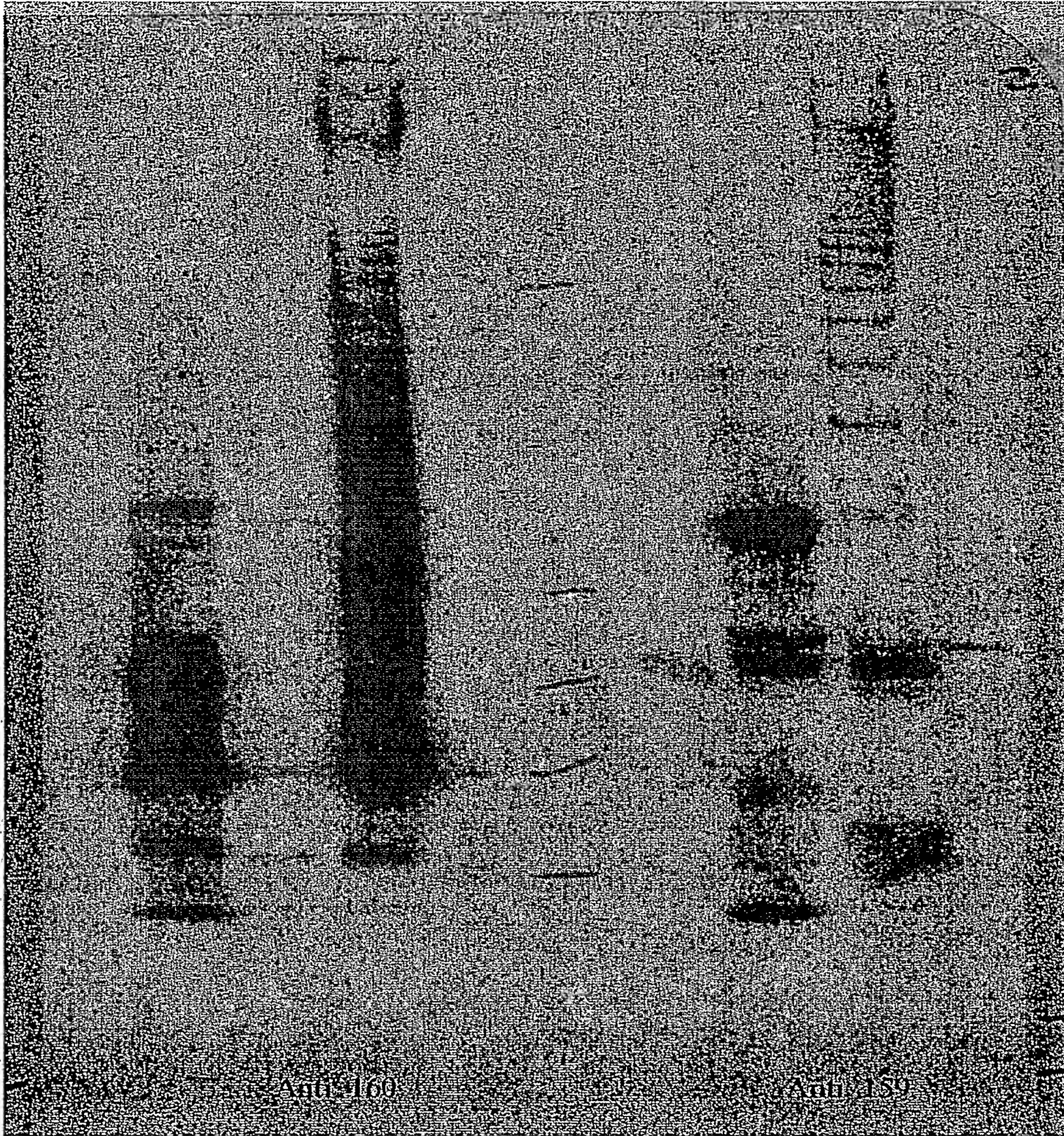


M6 strain isolate 2724

Figure 110

PCT/US05/27239

0160
0159
M6 3650 surf prot
M1 SF370 surf prot
Marker
0160
0159
M6 3650 surf prot
M1 SF370 surf prot



M6 strain isolate 3650

Figure 111

PCT/US05/27239

M6 2724 fraz art

M6 2724 tot

0160

Marker

158

0160

0159

M6 2724 surf prot

M6 2724 tot

M1 SF370 surf prot

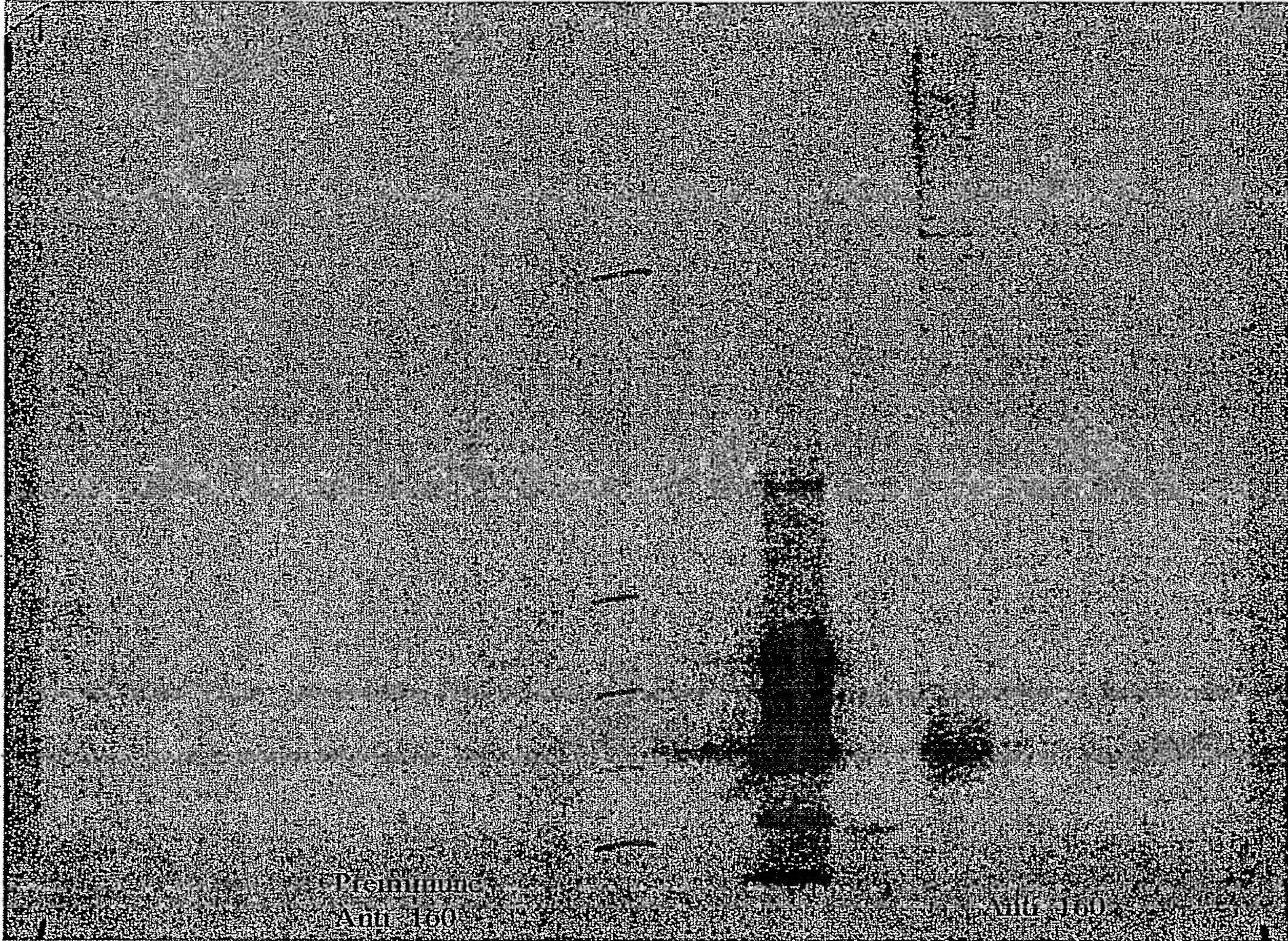


Figure 112

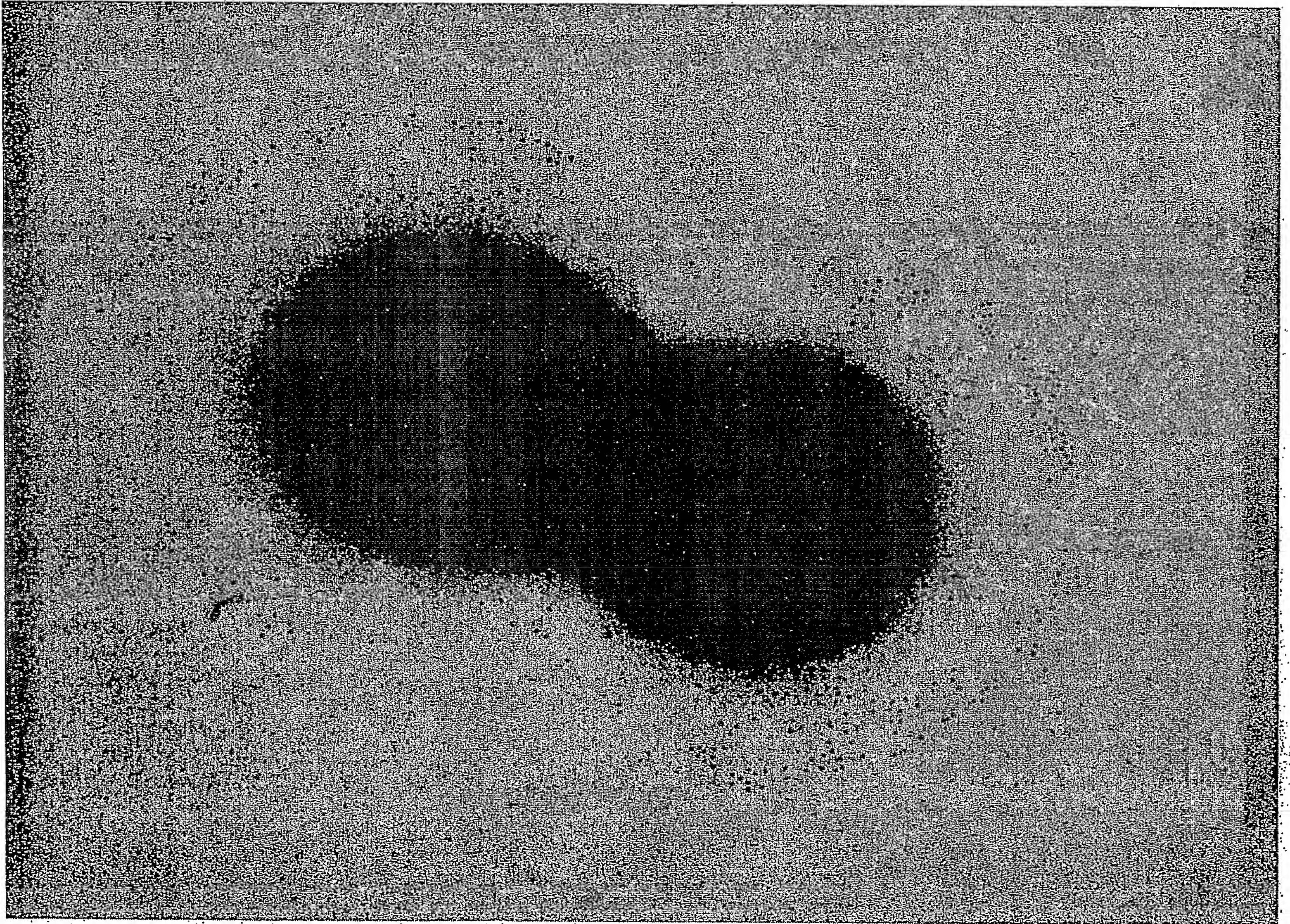


FIGURE 113

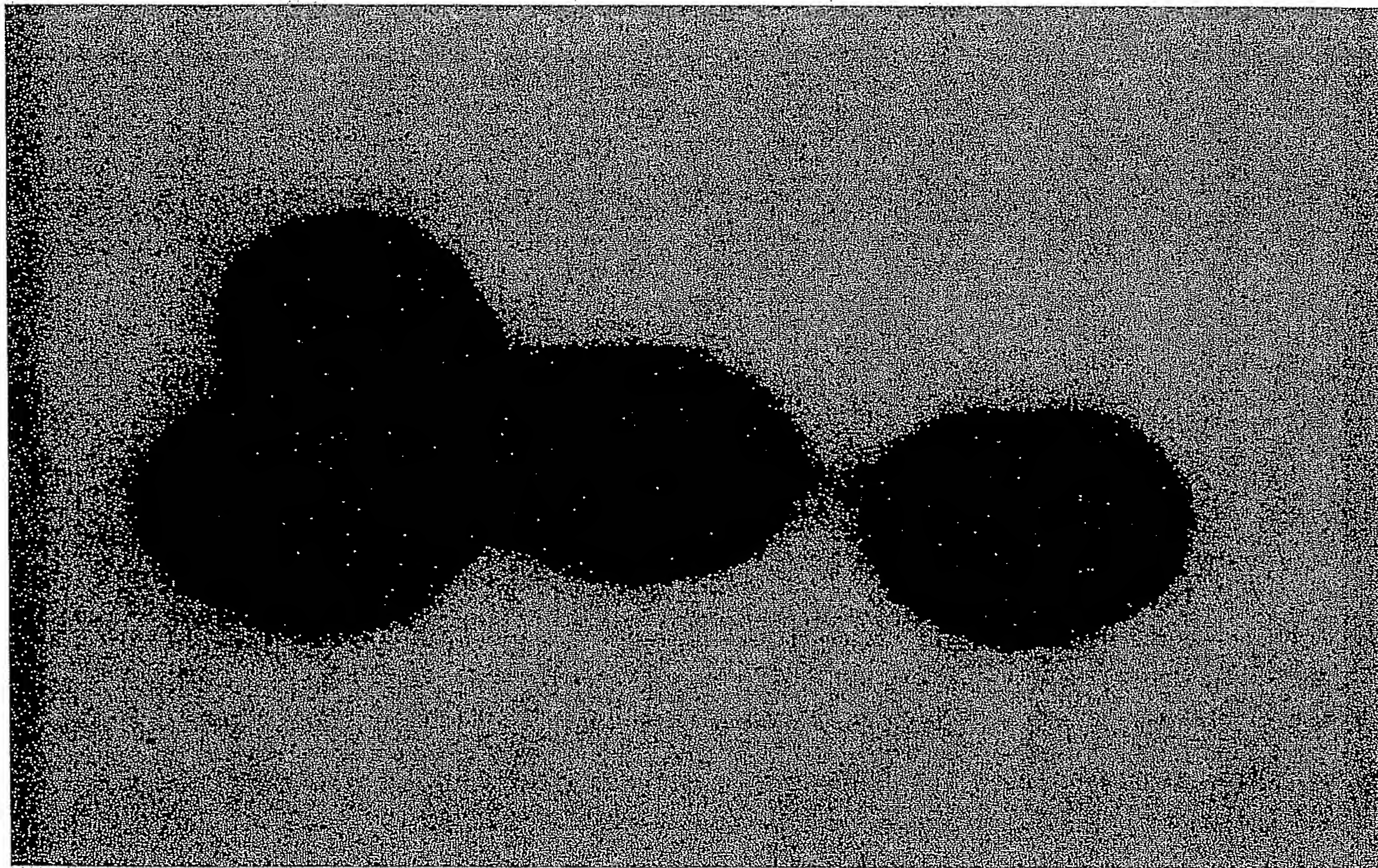


FIGURE 114

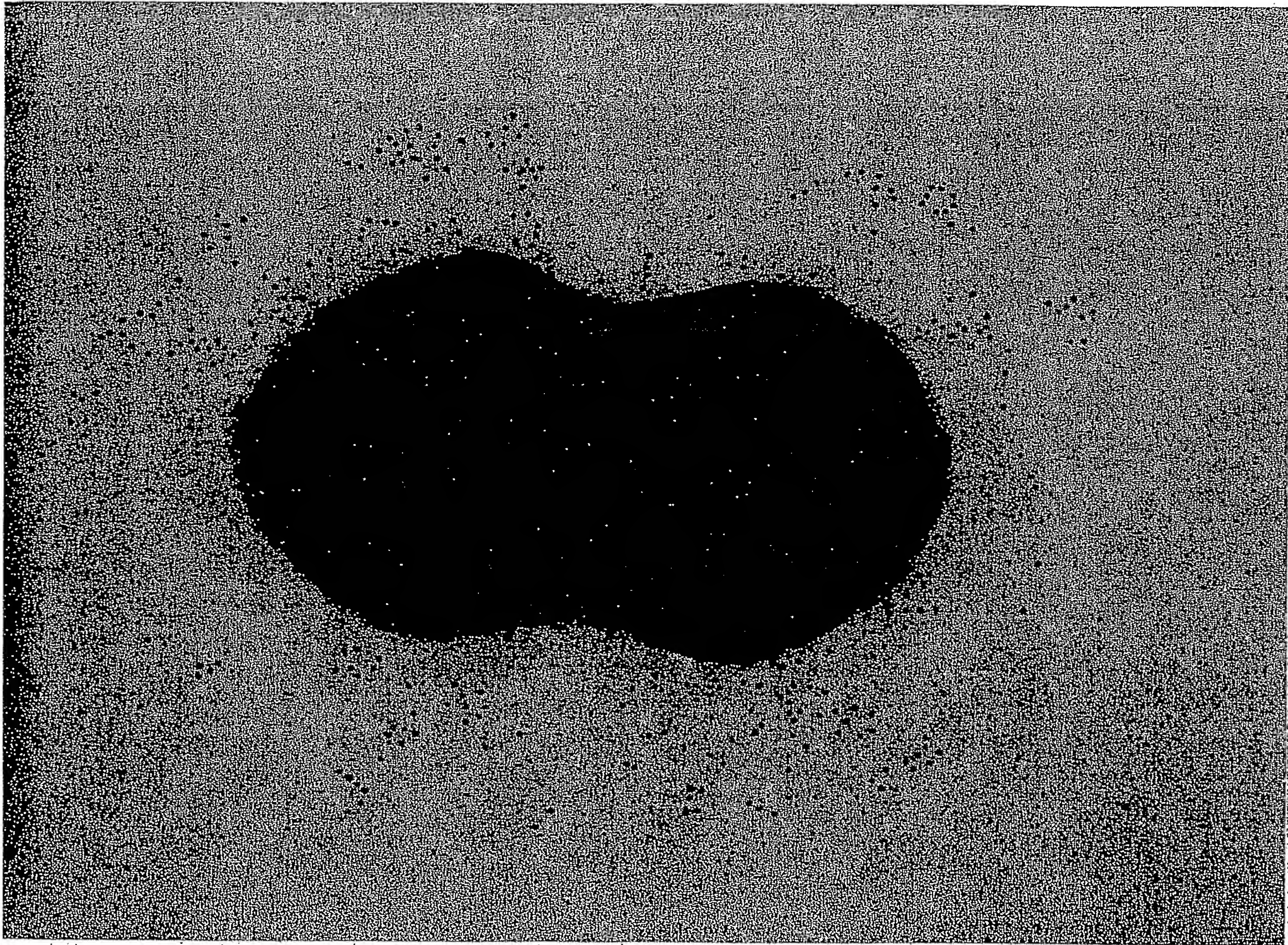


FIGURE 115

PCT/US05/27239 307/487

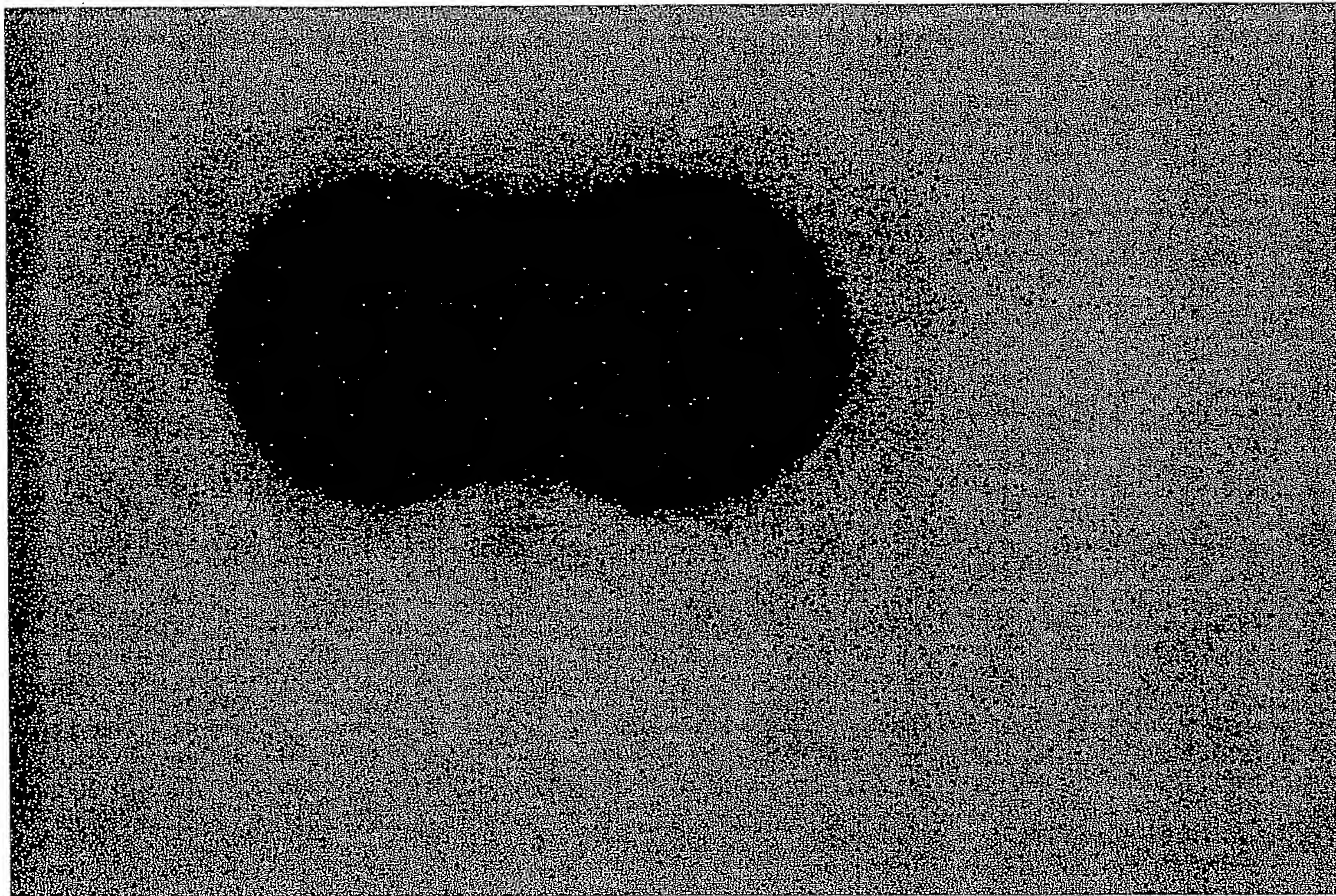


FIGURE 116

PCT/US05/27239

302/487

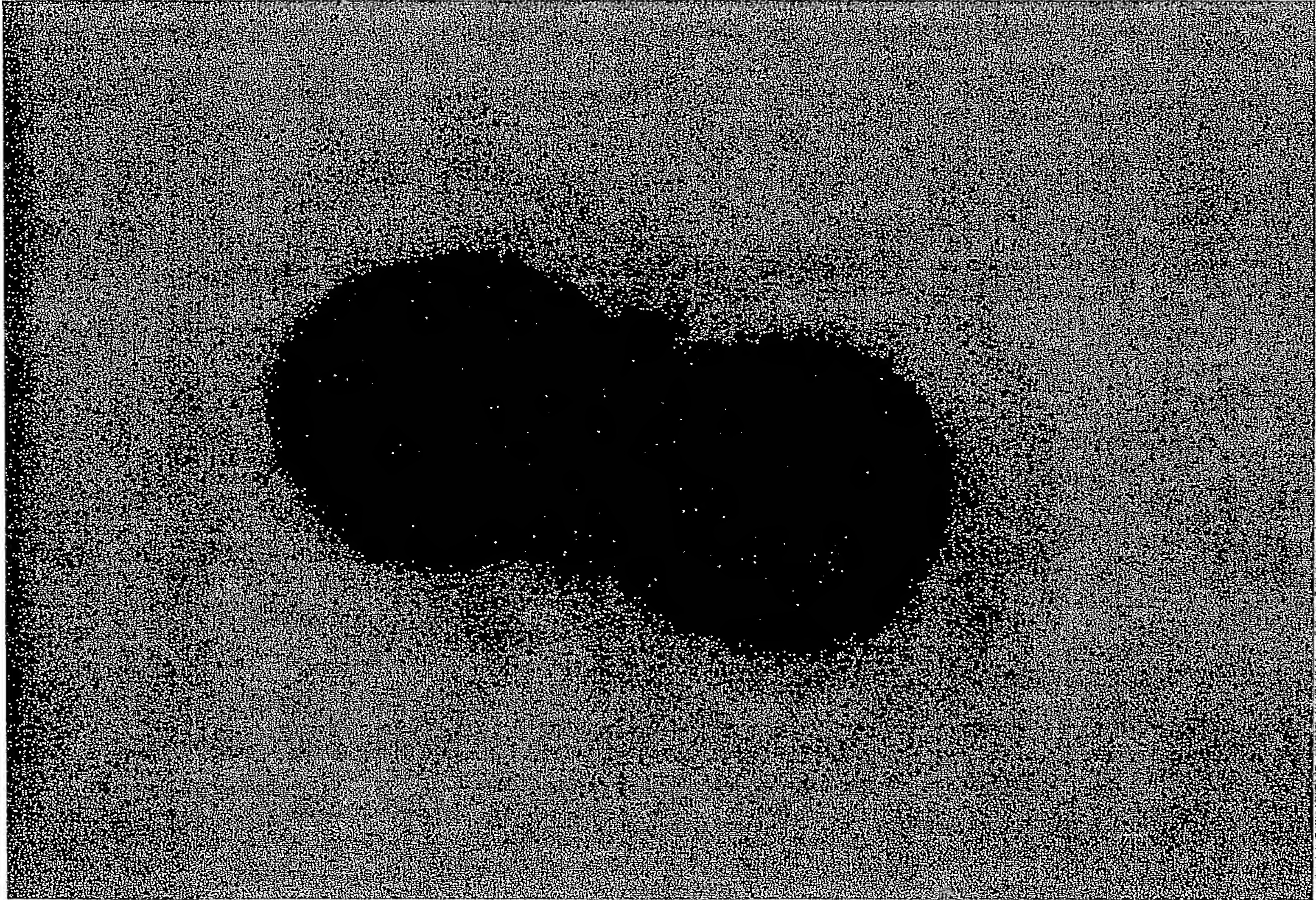


FIGURE 117

PCT/US05/27239 309/487

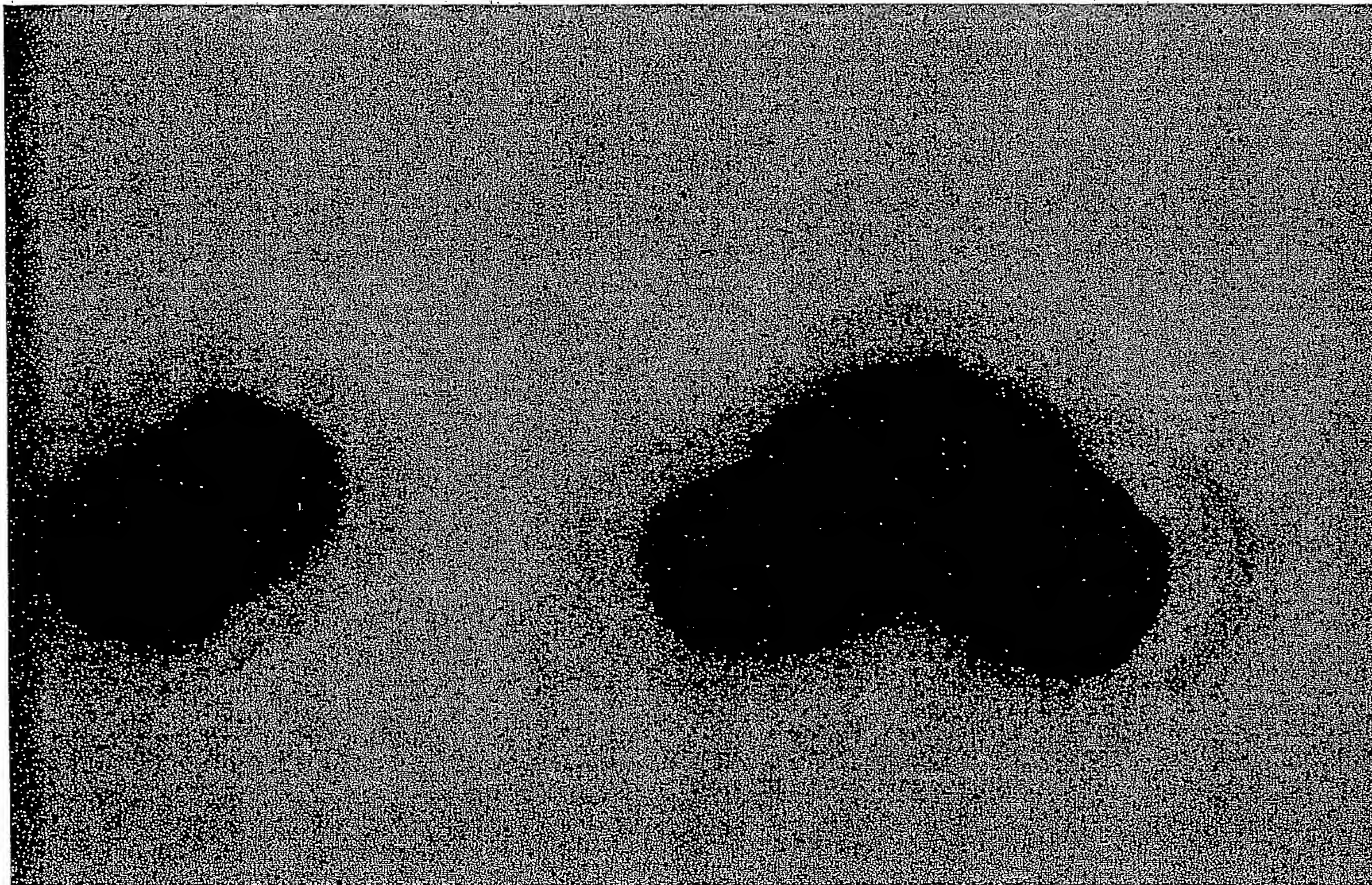


FIGURE 118

PCT/US05/27239 310/487

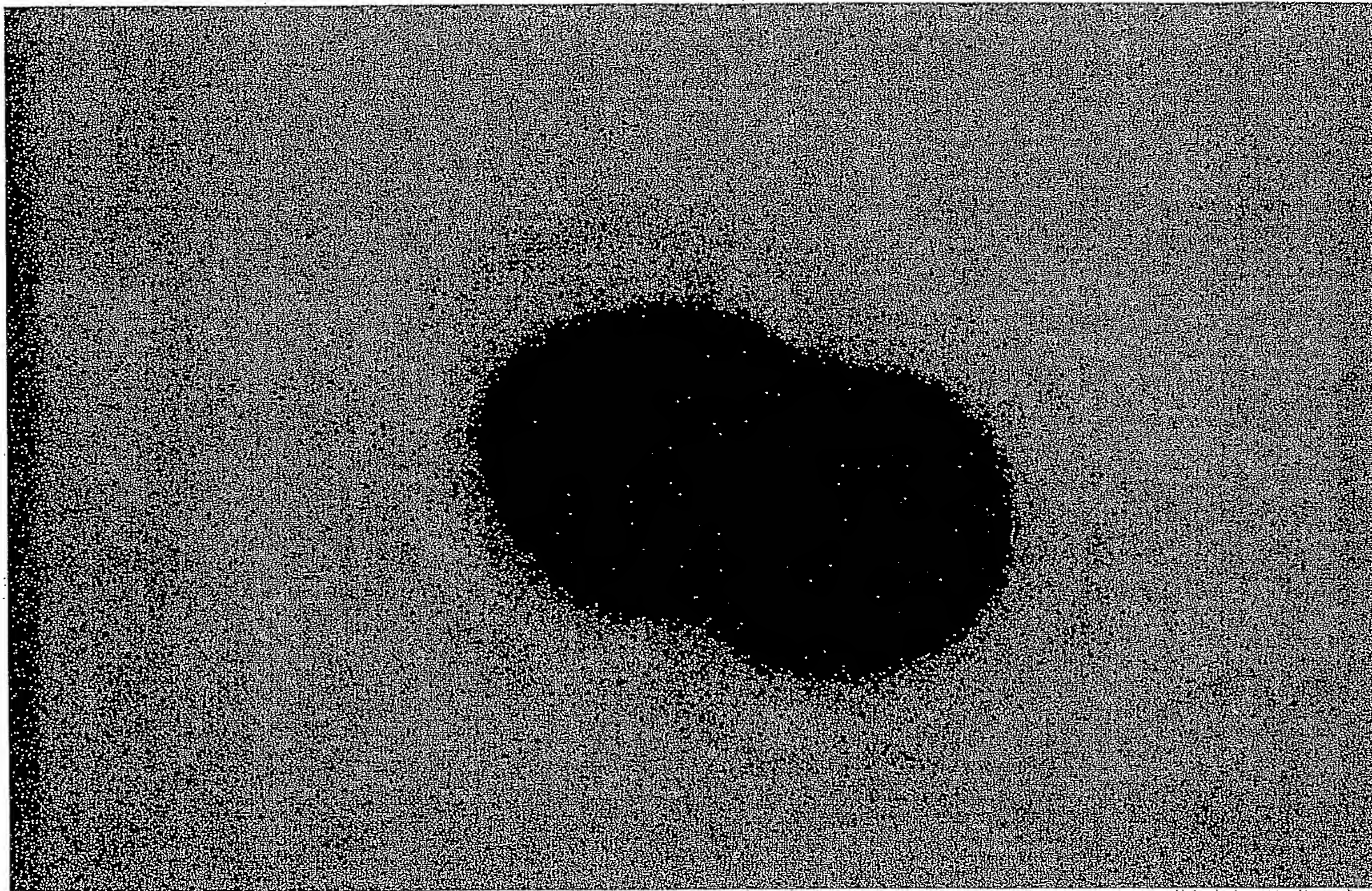


FIGURE 119

311/487

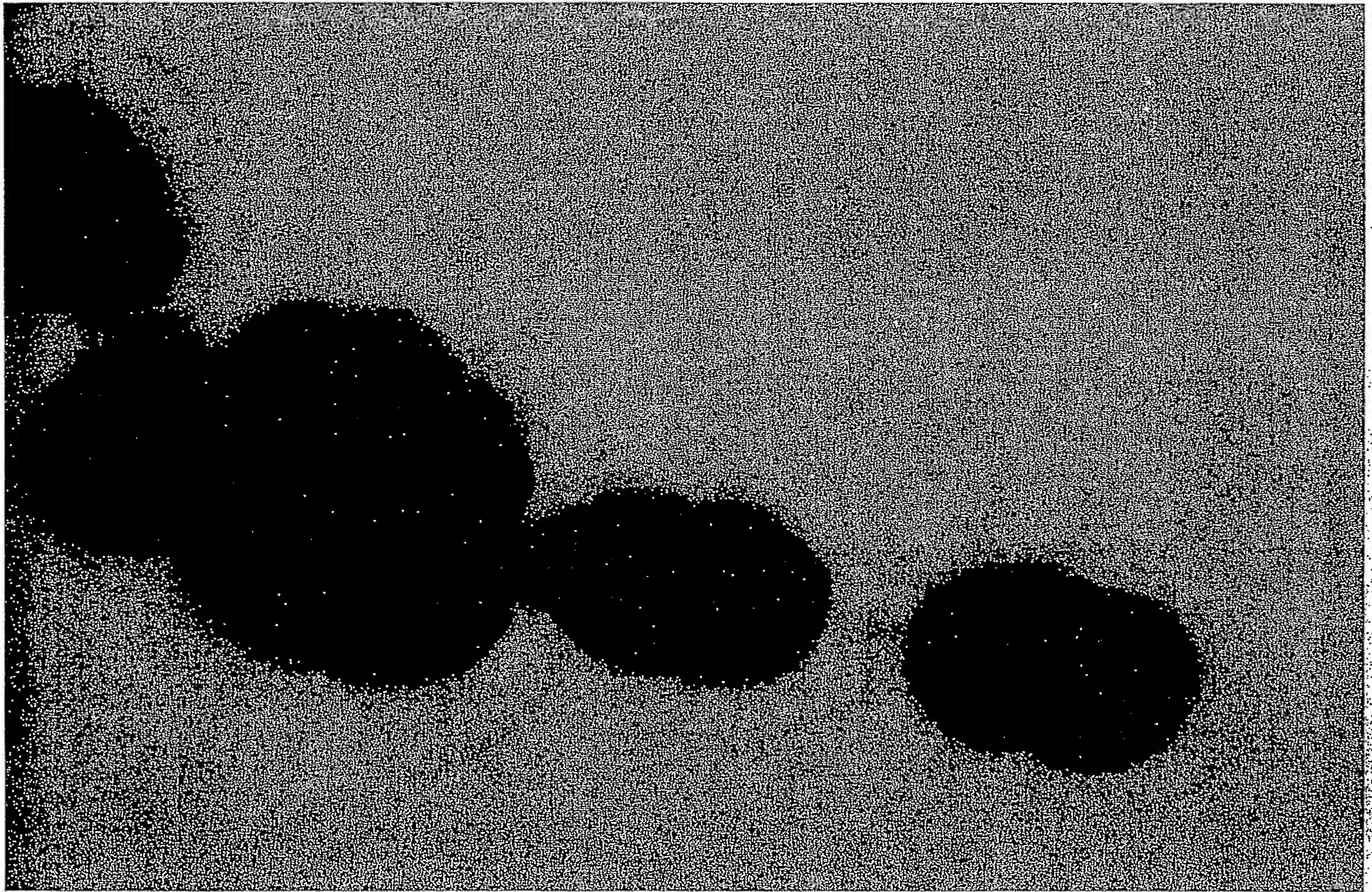


FIGURE 120

312/487

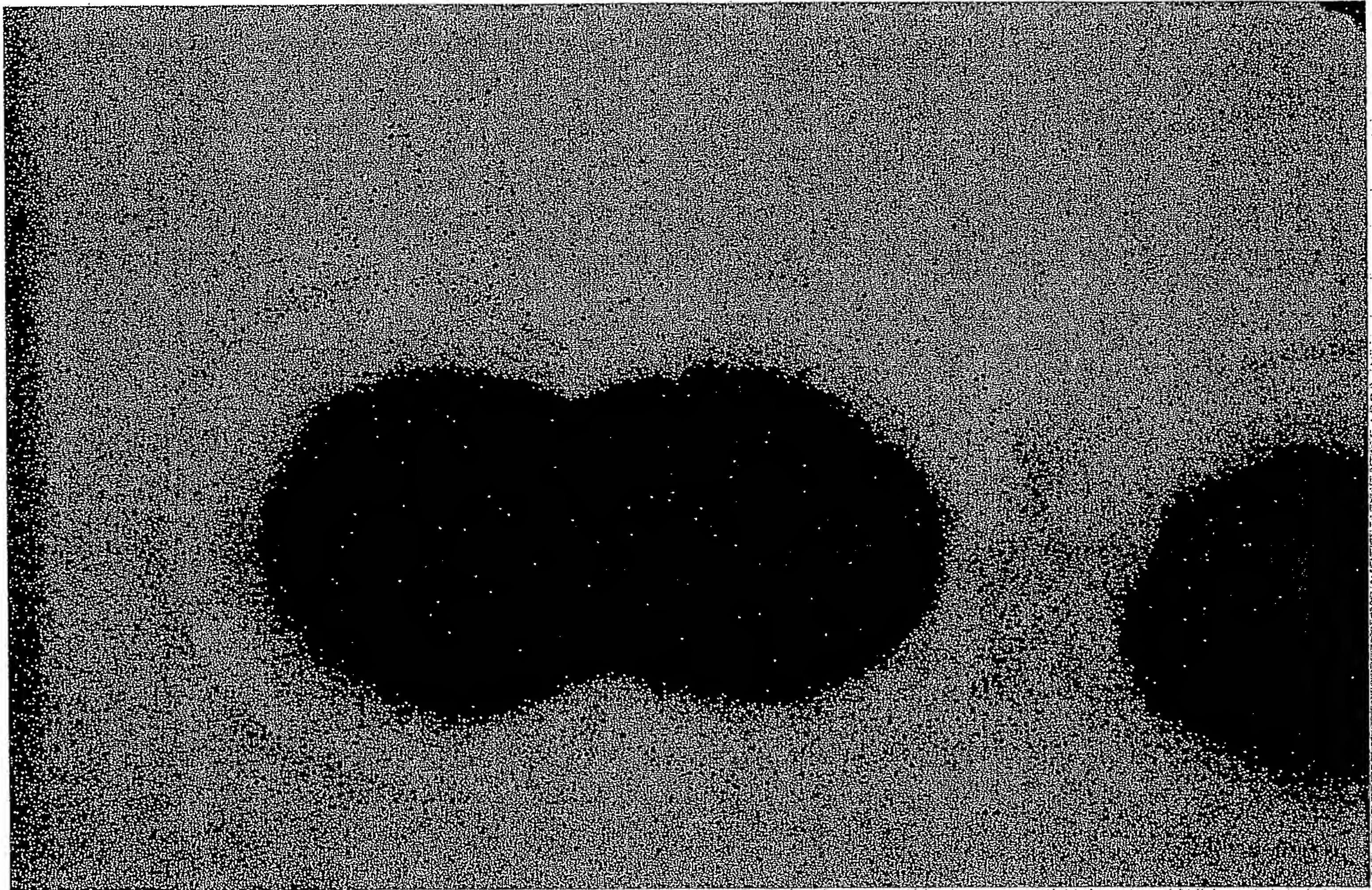


FIGURE 121

PCT/US05/27239

313/487

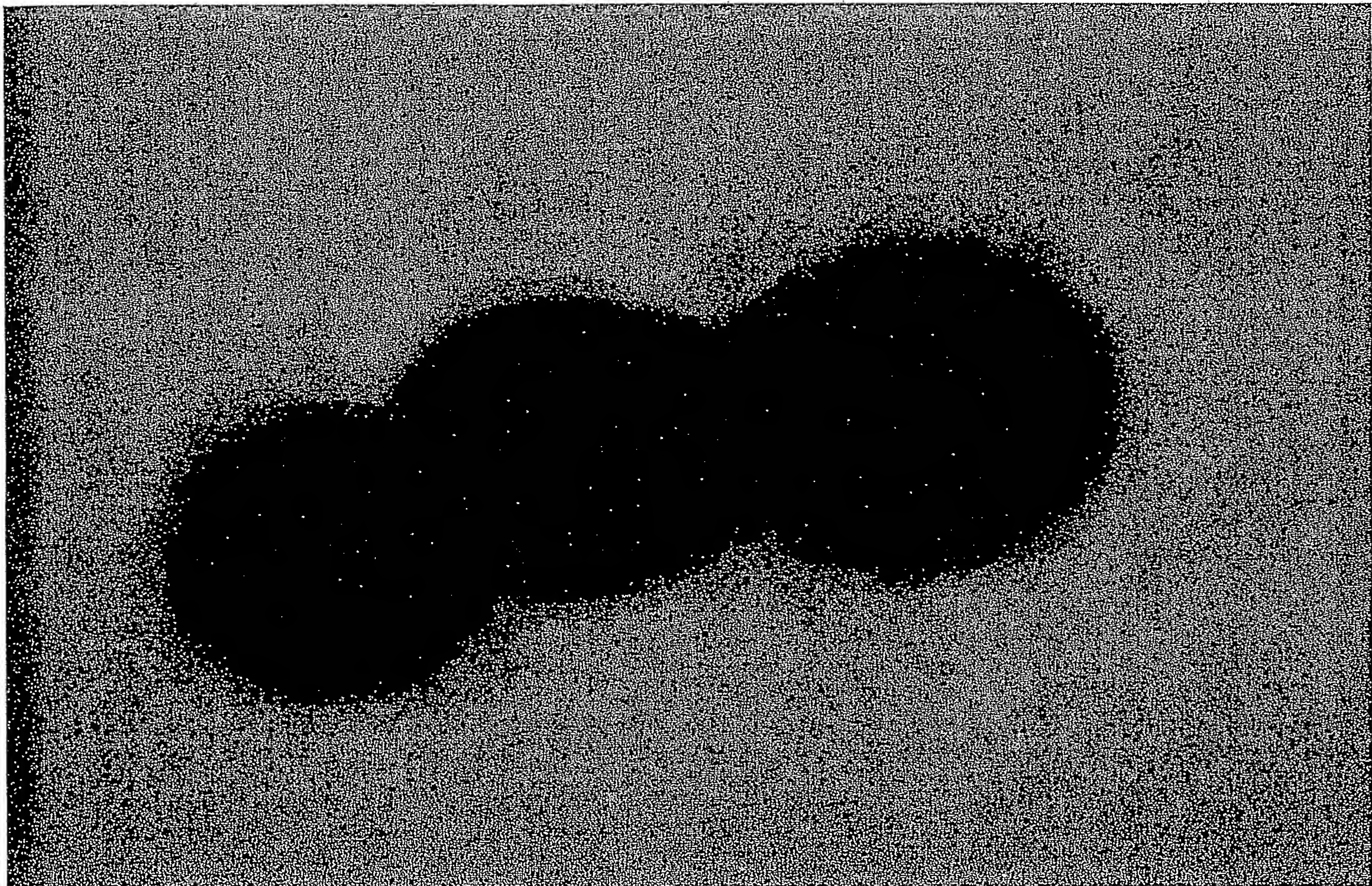


FIGURE 122

314/487

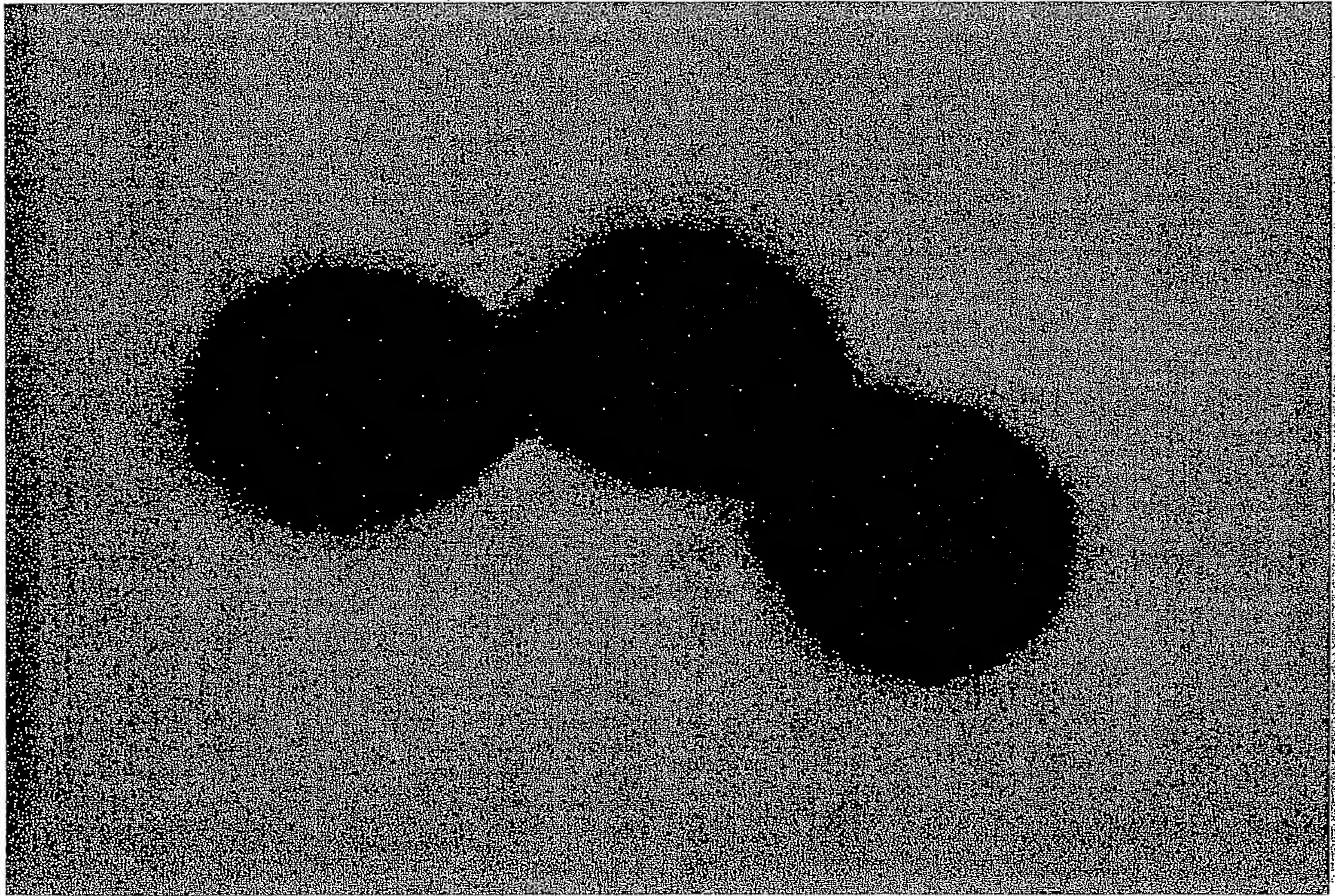


FIGURE 123

315/487

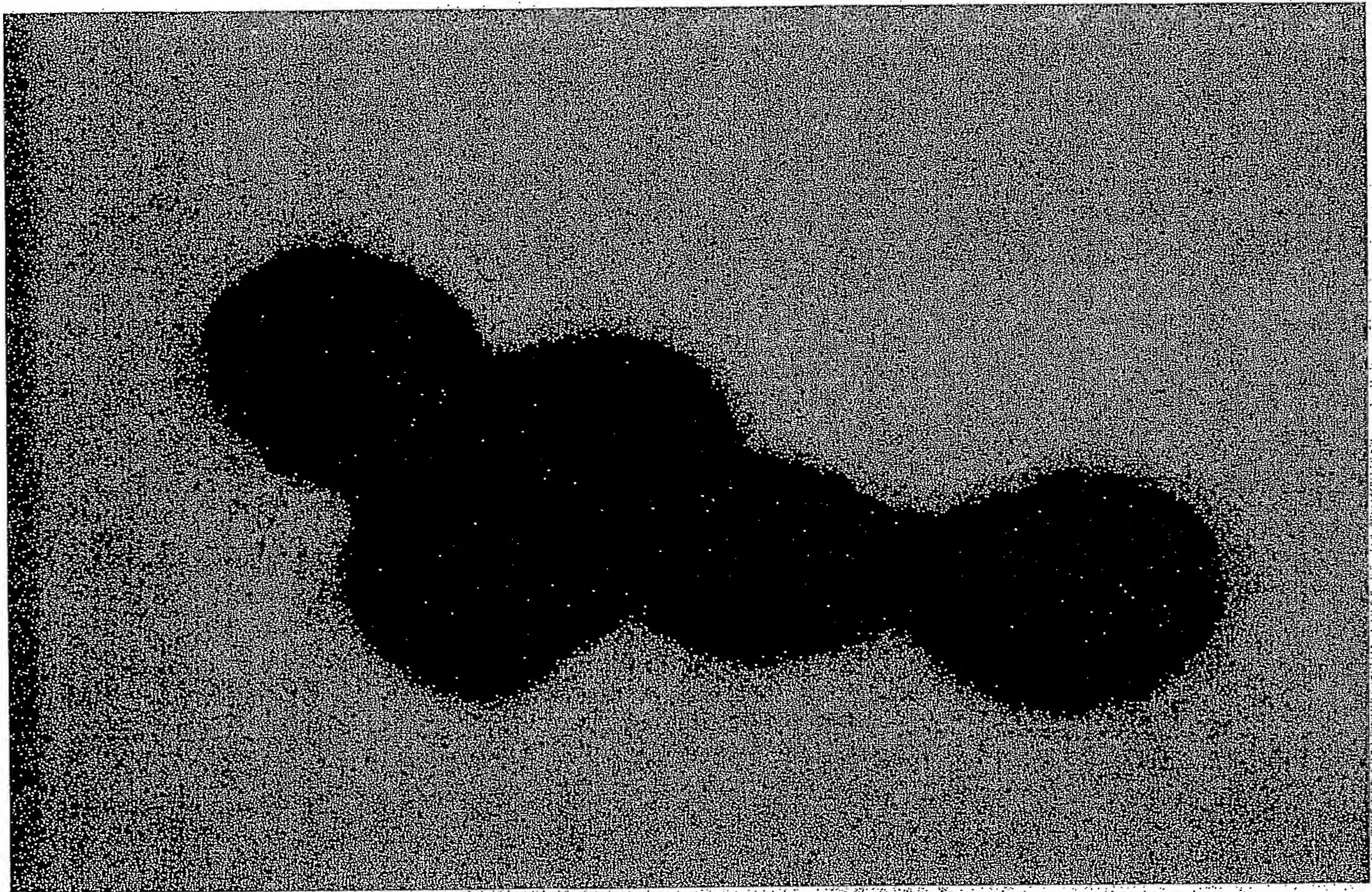


FIGURE 124

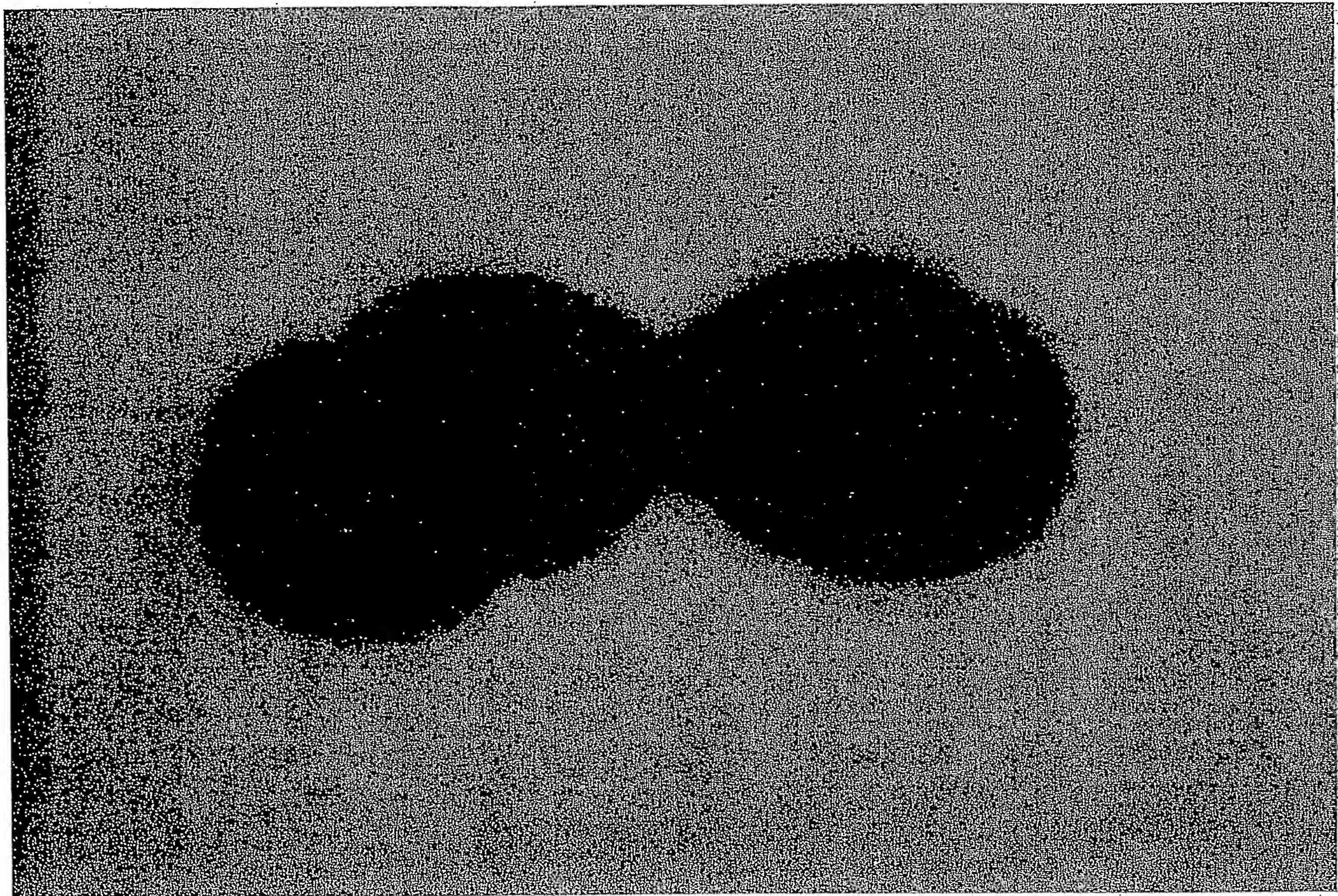


FIGURE 125

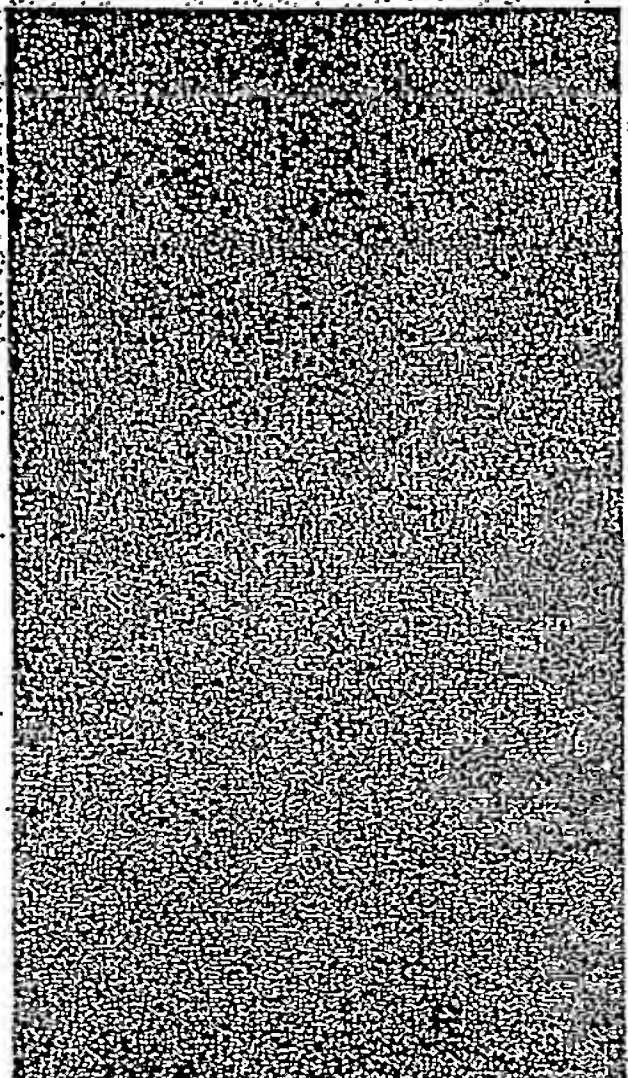
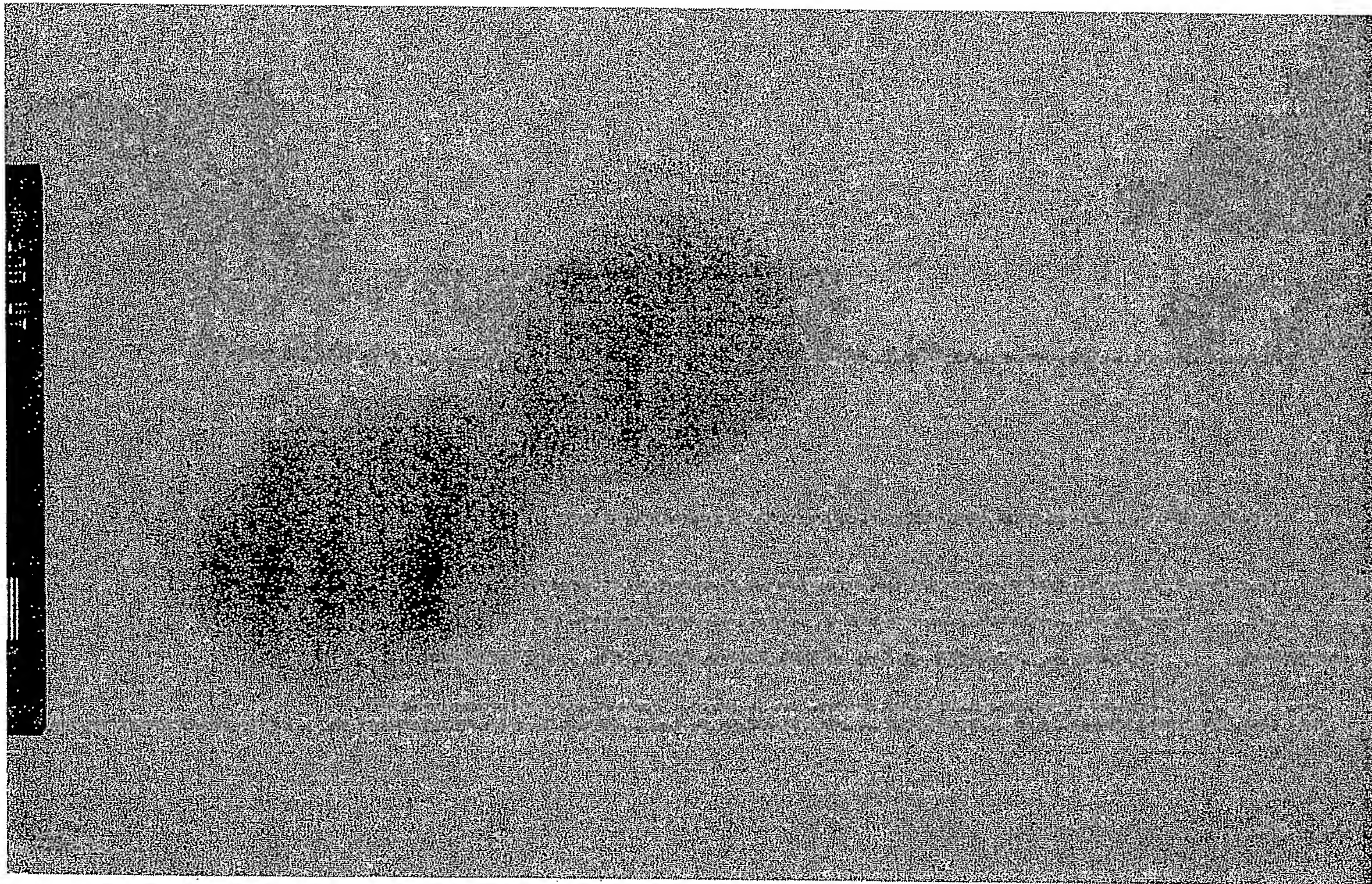


FIGURE 126

Figure 127



PCT/US05/27239 319/487

Figure 128

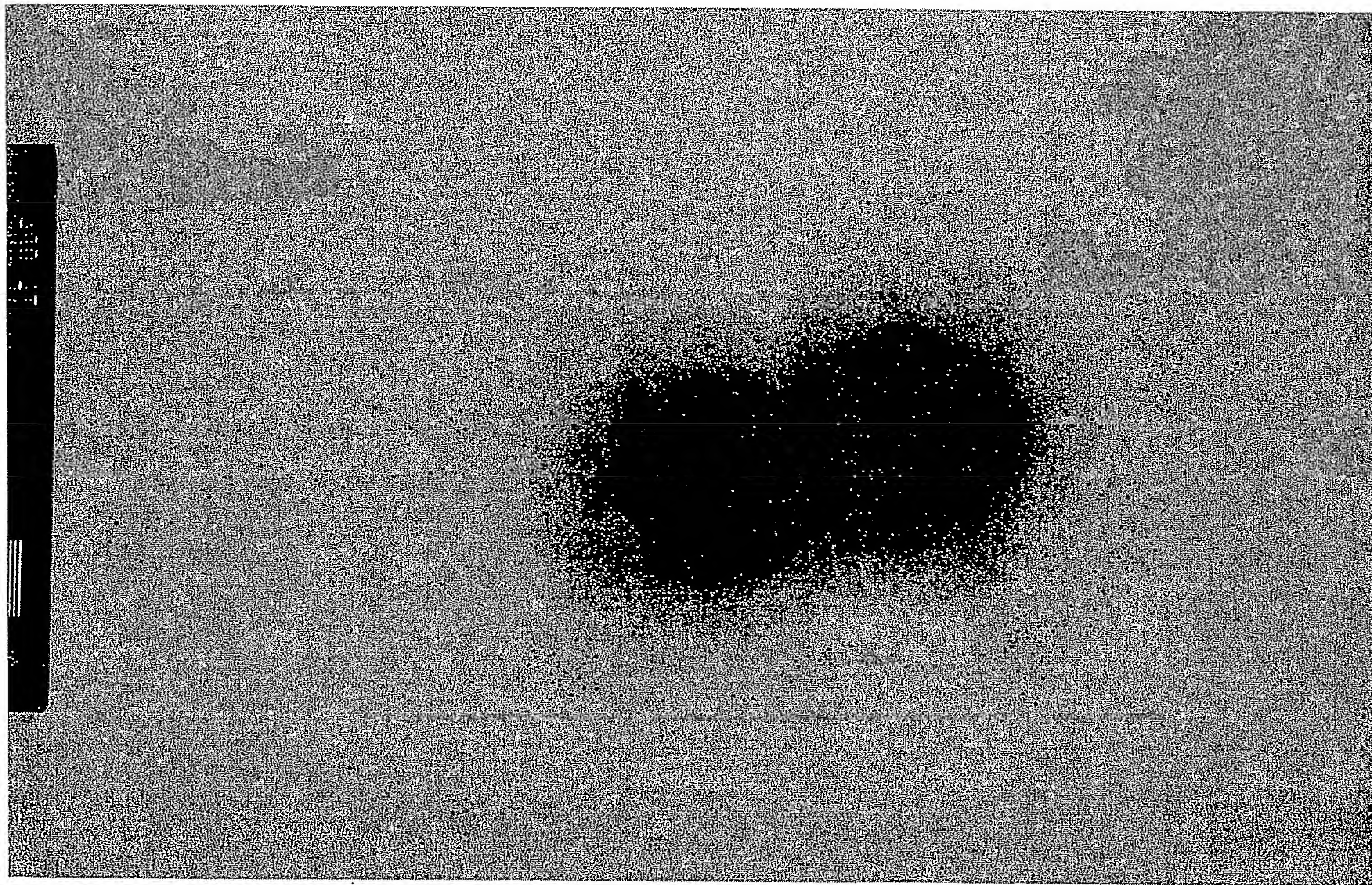


Figure 129

